Pervasive Technology Institute Annual Report: Research Innovations and Advanced Cyberinfrastructure Services in Support of IU Strategic Goals During FY 2016

Craig A. Stewart
Beth Plale
Von Welch
Geoffrey C. Fox
Matthew R. Link
Therese Miller
Eric A. Wernert
Michael J. Boyles
Ben Fulton
David Y. Hancock
Robert Henschel
Scott A. Michael
Marlon Pierce
Robert J. Ping
Tassie Gniady
Gary Miksik
Winona Snapp-Childs

Indiana University

PTI Technical Report PTI-TR16-007

Last revised December 15, 2016

Citation:

Stewart, C.A., Plale, B., Welch, V., Fox, G.C., Link, M., Miller, T., Wernert, E.A., Boyles M.J., Fulton, B., Hancock, D.Y., Henschel, R., Link, M.R., Michael, S.A., Pierce, M., Ping, R.J., Gniady, T., Mikski, G., & Snapp-Childs, W. (2016). Pervasive Technology Institute annual report: Research innovations and advanced cyberinfrastructure services in support of IU strategic goals during FY 2016 (PTI-TR16-007). Retrieved from http://hdl.handle.net/2022/21162.

PERVASIVE TECHNOLOGY INSTITUTE
INDIANA UNIVERSITY

Table of Contents

Table of Contents i

1. Executive Summary 1

2. Introduction 2

3. PTI Goal: Cultivate and enable creativity and innovation in science and scholarship by developing new innovations in cyberinfrastructure, informatics, and computer science 3

3.1. Center for Applied Cybersecurity Research (CACR) 3

3.2. Data to Insight Center 8

3.3. HathiTrust Research Center 9

3.4. Digital Science Center 10

3.5. National Center for Genome Analysis Support (NCGAS) 11

3.6. Research Technologies 14

4. PTI Goal: Create capabilities with which researchers at IU (and beyond) associate and collaborate 16

4.1. PTI Sub-goal: Be the 'partner of choice' within IU and the nation for creating and implementing cyberinfrastructure facilities 16

4.2. PTI Sub-goal: Enable the translation of software innovations to practical use 29

5. PTI Goal: Impact the economic health and quality of life in Indiana – creating new jobs, nurturing new businesses 31

6. PTI Goal: Support the development of a 21st century workforce within the State of Indiana 35

7. PTI Goal: Offer services that enhance enable new innovations and accelerate research by the IU research, scholarly, clinical, engineering, and artistic communities 40

8. Appendix 1: Listing of PTI and Research Technologies Facilities and Services 54

9. Appendix 2: EOT activities 62

1. Executive Summary

In 1999, under the leadership of Michael A. McRobbie (then Vice President of Information Technology), Indiana University proposed a partnership with Lilly Endowment to fund key elements of the launch of the School of Informatics and Computing and the creation of six advanced information technology labs that became known as the Pervasive Technology Labs (PTL). The core rationale had three main components: Indiana's economy was lagging behind many other states and in fact leading the nation in several negative economic indicators; state leadership had established a strategy of restoring the Indiana economy on the basis of life sciences and information technology; and, with funding to build on and expand existing strengths in computer science and information technology, IU could contribute strongly to economic as well as scientific and societal good in Indiana.

In 2008, the Pervasive Technology Labs into the Pervasive Technology Institute, enabled by a second round of funding from the Lilly Endowment and increased IU support. At the end of fiscal year 2016, the IU Pervasive Technology Institute (PTI) completed 17 years of innovation and service and completed its second year of sustainability without any funding from the Lilly Endowment. Continuing and refining its activities within IU, the state of Indiana, and the nation, some of the key accomplishments are:

- **CACR secures first "NSF Cybersecurity Center of Excellence" title.** In 2016, CACR secured a $5 million collaborative grant to lead the first NSF cybersecurity center of excellence (CCoE).
- **Jetstream accepted.** The Jetstream, a first-of-a-kind interactive cloud computing resource to support science and engineering research was accepted as a production system by the National Science Foundation.
- **National Center for Genomic Analysis supported.** The National Center for Genome Analysis Support (NCGAS) at Indiana University has received sustaining support from the National Science Foundation (NSF) to continue its work helping scientists analyze, understand, and make use of the vast quantities of genomic information.
- **Programmable Immersive Peripheral Environment Systems (PIPES) receives prestigious award.** PIPES is a tool that extends commonly used virtual reality systems in support cyber-physical applications – it was developed by Chauncey Frend from the Advanced Visualization Lab (AVL) and was awarded the "Best Research Demo Award" by the IEEE Computer Society (2016 IEEE Virtual Reality conference in Greenville, SC).
- **Project Innovation improves emergency response.** A new collaboration between the Bloomington Fire Department, Research Technologies, and software company Netage uses real-time data to help local firefighters work quickly and with the best possible information to help visualize 911 fire calls.
- **Advanced medicine supported.** The Structural Protein Ligand Interactome (SPLInter), supported by the Open Science Grid, is a computational drug design and discovery resource for ranking molecules docked to the human proteome. The overall impact of this project is that it is possible to identify small molecule candidates for individual proteins, or new protein targets for existing FDA-approved drugs and biologically active compounds.
- **SEAGrid Science Gateway aligned with Apache Foundation.** SEAGrid (formerly GridChem) science gateway, supported by SciGaP (NSF Grant #1339774), provides a platform for research and teaching in science and engineering and is now uses the Apache Foundation model for governance.
- **D2I plays key role in Midwest Big Data Hub**. IU is partnering with the University of Illinois, University of Michigan, University of North Dakota, and Iowa State University to in the Midwest Big Data Hub (NSF Grant #1550320) to improve the impact of data science.

- **HathiTrust Digital Library Corpus available for computational analysis**. During this year, through persistence and support from IU and D2I, HTRC overcame the monumental technical, legal, and architectural barriers to ingesting and securely provisioning access to the entire HathiTrust (HT) Digital Library Corpus (over 14 million texts and 5 billion pages) for computational analysis

2. Introduction

The Indiana University Pervasive Technology Institute's goals are to:

- Cultivate and enable creativity and innovation in science and scholarship by developing new innovations in cyberinfrastructure, informatics, and computer science
- Create association and collaboration capabilities for researchers at IU (and beyond) to:
    - Be the partner of choice within IU and the nation for creating and implementing cyberinfrastructure facilities (particularly when funded by a grant or contract focused on construction of a new facility or delivery of a new capability)
    - Enable the translation of software innovations to practical use
- Impact the economic health and quality of life in Indiana, creating new jobs and nurturing new businesses
- Support the development of a 21st century workforce within the state of Indiana
- Offer services that enable new innovations and accelerate research by the IU scientific, scholarly, clinical, engineering, and artistic communities

PTI is an inherently collaborative organization. Reporting organizationally within the Office of the Vice President for Information Technology (OVPIT), PTI is an institute created by collaborations across OVPIT, University Information Technology Services (UITS), the Maurer School of Law, and the College. PTI is organized into two types of functional units: research centers and service centers. The goal of this structure is to assure that collaboration, which has become a defining characteristic of PTI, crosses organizational boundaries, so practice can inform science and science can advance practice. PTI's research centers operate like any traditional research group. Their mission is to innovate within the scope of IU's mission The PTI research centers are:

- Center for Applied Cybersecurity Research (CACR), led by Von Welch
- Data to Insight Center, led by Professor of Informatics Dr. Beth Plale, also PTI Science Director
- Digital Science Center, led by Distinguished Professor of Computer Science Dr. Geoffrey C. Fox

The PTI Cyberinfrastructure[1] and Service centers are:

- Research Technologies Division of UITS, led by Associate Dean Dr. Craig A. Stewart, also Executive Director of PTI and adjunct associate professor of computer science in the IU School of Informatics and Computing
- The National Center for Genome Analysis Support, led by Dr. Thomas G. Doak

PTI operates within and supports Indiana University's vision (strategicplan.iu.edu):

> *To be one of the great research universities of the twenty-first century and to be the preeminent institution of higher education in Indiana, specifically by:*

[1] Cyberinfrastructure "consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked by high speed networks to make possible scholarly innovation and discoveries and new artistic expressions not otherwise possible" as defined in Stewart et al. 2014 https://scholarworks.iu.edu/dspace/handle/2022/18608.

1. Providing an excellent, relevant and responsive education across a wide range of disciplines in baccalaureate, graduate, and professional education to students from all backgrounds from Indiana and around the globe.

2. Pursuing excellent world-class research, scholarship, and creative activity.

3. Engaging in the economic, social, civic, and cultural development of Indiana, the nation, and the world by building on the base of excellence in research and education.

3. PTI Goal: Cultivate and enable creativity and innovation in science and scholarship by developing new innovations in cyberinfrastructure, informatics, and computer science

PTI overall and particularly the PTI Research Centers, which operate as small and nimble R&D centers, lead PTI in development of new innovations – and are the primary elements of PTI developing new innovations in cyberinfrastructure (CI), informatics, and computer science. Metrics describing research and innovation achievements during FY 2016 are summarized in Table 3-A and then described in more detail for each center below.

Table 3-A. Research and innovation productivity metrics for PTI from inception of PTL in 1999

	Publications	Technical Presentations	Nobel Prize Awards Supported	Open Source Software Released
PTL & IUPUI 1999-2014	1,796	1,426 (data from 2008 forward)	3	178
PTI FY 2015	39	15	0	10
PTI FY 2016	34	55	0	8
Totals	**1,835**	**1,441**	**3**	**A total of 10 current open source products are released by PTI**

3.1. Center for Applied Cybersecurity Research (CACR)

In 2016, CACR secured a $5 million collaborative grant to lead the first NSF cybersecurity center of excellence (CCoE). Building on CACR's success with CTSC, the Center for Trustworthy Scientific Cyberinfrastructure (CTSC), the CCoE is comprised of cybersecurity experts who have spent decades working with science and engineering communities and have an established track record of usable, high-quality solutions suited to the needs of those communities. The team draws from best operational practices and includes leaders in the research and development of new methodologies and high-quality implementations.

The mission of the NSF CCoE is to improve the cybersecurity of NSF science and engineering projects, while allowing those projects to focus on their science endeavors.

This mission is accomplished through one-on-one engagements with projects to address their specific challenges; education, outreach, and training to raise the state of security practice across the scientific enterprise; and leadership on bringing the best and most relevant cybersecurity research to bear on the NSF cyberinfrastructure research community. Additionally, the CCoE hosts the annual NSF Cybersecurity Summit, bringing the NSF and research communities together to build understanding of the information assets that enable science, while providing the community a forum for education, sharing experiences, building relationships, and establishing best practices.

Highlights of the CCoE (www.trustedci.org) accomplishments since it's inception in January 2016 include:

- Engagements with four NSF projects:
    - **Gemini Observatory** executed an extended engagement with CCoE focused on core policy processes and documentation, as well as a close unified look at technical, and physical controls at Gemini North.
    - **Image Based Ecological Information System (IBEIS)** is developing a software platform to collect and share animal data for science and conservation. CCoE collaborated with IBEIS staff to prototype role-based access control for the IBEIS platform.
    - The **Array of Things (AoT)** project is, in collaboration with the City of Chicago, developing and deploying a network of interactive, modular sensor boxes that will be installed around Chicago. CCoE completed an assessment of AOT's cybersecurity, advised AoT and the City of Chicago's CIO on best practices for developing privacy policy, and assisted them in processing the feedback they received on their draft privacy policy from privacy advocates and Chicago residents.
    - **SciGap** is developing a set of core infrastructure services to support science gateways. CCoE and SciGaP collaborated to design the security and identity management functionality of those services.
- In collaboration with ESnet, convened a working group of leaders from the open science community to develop an Open Science Cyber Threat Profile.
- Initiated a Situational Awareness service for the NSF community.
- Launched the CCoE Webinar Series first two webinars drawing over 30 total attendees from the community.
- We developed and presented a short one-hour version of our cybersecurity program training targeted to small to medium sized NSF projects at Indiana University.

Highlight: CACR and Crane collaborate on cybersecurity efforts

Signed at the end of FY2016, and announced in July, CACR entered into a two-year collaborative agreement with the Naval Surface Warfare Center, Crane Division. This agreement, a Cooperative Research and Development Agreement (CRADA), brings two of Indiana's leaders in cybersecurity together in collaboration to share personnel and expertise to advance research and development in tackling cybersecurity challenges to our Nation.

Captain Jeffrey T. Elder, Commanding Officer of NSWC Crane (left) with VPIT Brad Wheeler (right) signing cybersecurity agreement.

3.1.1. CACR expands community and training opportunities around NSF cyberinfrastructure ecosystem.

Through the Center's leadership of the Center for Trustworthy Scientific Cyberinfrastructure (CTSC), CACR has planned and executed the NSF Cybersecurity Summits for Large Facilities and Cyberinfrastructure in 2015 as it has for the past three years. The NSF cyberinfrastructure ecosystem presents an aggregate of complex cybersecurity needs (e.g., scientific data and instruments, unique computational and storage resources, complex collaborations) as compared to other organizations and sectors. This community has a unique opportunity to develop information security practices tailored to these needs, as well as break new ground on efficient, effective ways to protect information assets while supporting science. With ninety attendees from 50 NSF funded programs, 14 being Large Facilities, the Summit brings together leaders in NSF cyberinfrastructure and cybersecurity. This allows those in attendance to annually work toward building a trusting, collaborative community, and seriously address the community's core cybersecurity challenges. Viewed as critical to cybersecurity, at least one NSF solicitation has made attendance a mandatory component of the award.

3.1.2. CACR sees continued community success with annual Cybersecurity Summit.

CACR has been bringing together leading visionaries in the area of applied cybersecurity technology, education, and policy in an annual Cybersecurity Summit since 2010. During this one-day event, attendees discuss the proper balance of public needs, homeland security concerns, and individual privacy

rights. The 2015 CACR Cybersecurity Summit focused on *Privacy and Risk Management*, and featured keynotes from NIST's Ron Ross as well as Harvey Rishikof of the ABA Cybersecurity Legal Taskforce. There were 180 individuals in attendance, representing 73 organizations and institutions such as Raytheon, Indiana Department of Homeland Security, University of Kentucky, Indiana Office of Technology, Barnes & Thornburg, Purdue University, Eli Lilly and the Indiana Economic Development Corporation. Additional information about the Summit can be found at: https://cacr.iu.edu/events/cybersecurity-summit/index.php

3.1.3. CACR provides cybersecurity expertise to Open Science Grid.

The Open Science Grid (OSG) is a nationwide facility and infrastructure enabling large-scale high-throughput computing for science. PTI already provides the Grid Operations Center (GOC) support for the OSG and in FY 2016, CACR joined the collaboration, working alongside the GOC, contributing cybersecurity expertise as part of OSG's security team.

3.1.4. CACR helps secure patient health information end to end.

CACR launched its healthcare cybersecurity initiative this past year by assuming responsibility for HIPAA consulting services as part of its portfolio. Designed to help organizations secure patient health information end to end, the new services use an innovative, more comprehensive approach that weaves cybersecurity into user workflows. CACR leverages the widely used NIST risk management framework, which allows organization to address HIPAA and FISMA compliance simultaneously.

During the past year, CACR has provided welcome HIPAA guidance to national labs, supercomputer centers, and universities, but its largest impact has been at home. CACR's HIPAA consulting services raised healthcare cybersecurity across Indiana University by assisting University Information Technology Services achieve or maintain HIPAA compliance for central university systems that store protected health information for Indiana University Schools of Medicine, Dentistry, Nursing, and Optometry. These include systems managed by the Client Services and Support (8), Enterprise Systems (8), Learning Technologies (2), Networks (3), and Research Technologies (20) divisions, 41 in total. CACR also engaged in over 100 HIPAA consultations and collaborative projects with researchers and IT professionals within the Indiana Clinical and Translational Sciences Institute; trained over 350 University Information Technology Services staff on the HIPAA Security Rule, delivered a presentation on HIPAA compliance at the Statewide IT Conference; and offered a risk management course for Indiana University users at large.

3.1.5. CACR leads creation of cybersecurity.iu.edu.

CACR led the development of cybersecurity.iu.edu, a web site acting as a repository for information regarding cybersecurity at Indiana University. A collaborative effort between CACR, the School of Informatics and Computing, the Kelley School of Business, the Maurer School of Law, Public Safety and Institutional Assurance and REN-ISAC, this site is the university's comprehensive approach to exploring the new frontier of cybersecurity.

3.1.6. CACR Speaker Series draws experts from across the country.

Held at least monthly, CACR's Security Seminar Speaker Series presentations allow cybersecurity professionals from all over the country to give talks on their individual areas of expertise. These talks are open to all interested students, faculty and staff, and are offered at IUPUI via live stream. The 2015-2016 Speaker Series, attended by 262 individuals.

3.1.7. Outreach Efforts

CACR's mission is to advance the state of cybersecurity practice, interdisciplinary research, and understanding in order to serve Indiana University, the state of Indiana, and our national and global communities. One of the features of the Center's work is to focus on security challenges in context. This is done through outreach with the various communities the Center serves. In the past year, some of the most impactful outreach that has been offered has been:

- National Security Ethics Workshop: In November of 2015, CACR co-sponsored a visit from U.S. Army War College's Dr. Leonard Wong. Dr. Wong participated in an "Ethics for Breakfast" event with undergraduate students (45 attendees), a faculty workshop entitled "Ethics, Leadership, and National Security" (37 attendees) and a public lecture (4 attendees) entitled "Societal Trust in the Military."
- ERA Presentation: In May of 2016, CACR presented "Everyday Cybersecurity" to the Electronics Representatives Association of Indiana and Kentucky. This presentation, directed at small and medium business owners, provided the 20 attendees with easy to implement measures that would make their lives and businesses significantly more secure.
- IT Security Training for Lawyers: In December of 2015, CACR held an IT security training for lawyers at Barnes & Thornburg, entitled "Practical Cybersecurity for Lawyers and Law Firms." This training focused on growing the attendees' literacy and perspective around cybersecurity, providing the 46 attendees with specific, practical, doable guidance and helping attendees develop the know-how to get technical expertise when they need it.
- Practical Cybersecurity for Open Science Projects: Held in April of 2016, this CACR led training addressed for the 7 in attendance the information security requirements outlined in National Science Foundation Major Research Equipment and Facilities Construction / Large Facility cooperative agreements, and to provide guidance, tools, and resources for open science projects of all sizes.

3.1.8. CACR begins to focus on the next generation.

CACR held its inaugural Security Matters Cybercamp for high school students at the end of FY2016. Attended by 22 students from Bloomington, Indianapolis and New Albany, the camp was launched by CACR to address the need for cybersecurity outreach and education for K-12 students. The camp's purpose was to educate youth about the importance of online security and privacy matters, equip them with tools and knowledge to protect them against cybercrime and other cyberthreats, and allow them to see what being a cybersecurity professional can look like, all while addressing the need to increase the pipeline of young professionals into the field.

3.2. Data to Insight Center

The Data To Insight Center (D2I) engages in interdisciplinary research and education in the preservation of scientific data, Big Data, cyberinfrastructure, frameworks for distant reading in digital humanities, provenance, and cloud computing. It carries out data modeling for large and complex data, pioneers new forms of data publishing including tools for embedded data curation early in the lifecycle of data and models for persistent identifier user; innovates through new tools for provenance capture; pioneers novel Big Data cyberinfrastructure; studies communities of practice in data sharing, and develops early career data scientists through a fellows program with the Research Data Alliance.

In FY 2016, there were several milestones, captured in the project descriptions below.

3.2.1. SEAD

This year was a major milestone for the SEAD D2I team as it completely reworked its publishing services as part of the SEAD 2.0 release (July 2016). The SEAD 2.0 architecture gives the researcher and project a cloud hosted project space, and the D2I contributed a publishing pipeline from the project space to a choice of repositories. SEAD publishing services include:

- Curbee: a suite of microservices that are applied to research objects to enhance (curate) them for publication.
- SEAD Matchmaker: a recommendation engine that identifies most appropriate repositories for publication and publishes research objects.
- PDT (People, Data, Things): a repository of profiles about the people, data, and repositories that are used in SEAD
- IU SEAD Cloud: a reusable, thin preservation and repository layer over an HPC storage system. It is deployed at IU over an HPSS replicated tape store. IU SEAD Cloud serves as a repository for SEAD users through an agreement between co-PI Beth Plale and PTI Executive Director Craig Stewart to use IU Scholarly Data Archive in support of SEAD submissions. T

SEAD @ IU strengthened its partnership with the NSF-funded DataONE project through improved interoperability and metadata sharing between the projects.

3.2.2. Research Data Alliance (RDA)

D2I is involved in several initiatives with the Research Data Alliance (RDA), including building RDA presence in the US through leadership and engagement, advancing early career researchers and professionals in RDA through the Data Share fellowship program), and the RDA-PRAGMA adoption grant. The leadership activities during this year resulted in the development of volunteer value proposition for RDA/US and refinement of several focused projects, such as the RDA/US Curator initiative and persistent identifiers (PID) testing and training.

The RDA/US Curator project conceived as part of the RDA/US leadership activities attracted qualified early career professionals, one of whom then became a fellow in the Data Share program.

The RDA Data Share program, an early career fellows program funded by a grant to D2I by the Alfred P. Sloan Foundation, selected 14 fellows in 2016 – almost a doubled increase compared to the previous year. The 2016 class of fellows will attend their first RDA Plenary September 2016 in Denver, CO.

D2I, with collaborators in Japan and the Philippines, received an RDA/US Adoption Award through funds by the MacArthur Foundation. The Adoption Award is to fit two recent formal RDA recommendations (outputs) with persistent identifiers (PIDs) into the publishing process for the rice genomics analysis carried out on the Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) Cloud Testbed. The project, called PRAGMA Rice, involves integrating both PIDs and

provenance into the publishing process from PRAGMA VMs that can exist on any number of Pacific Rim clusters that make up the PRAGMA Cloud Testbed.

Through the MacArthur funding a D2I doctoral student spent 5 weeks as a visiting scholar in AIST (The National Institute of Advanced Industrial Science and Technology) in Tsukuba, Japan, where he continued to gather requirements and test the prototype for the project. Results of this work will be presented at the RDA plenary in Denver and at the PRAGMA meeting, both in September 2016.

D2I contributed to the summer school at Trieste Italy in August 2016. Devan Donaldson, Asst. Prof at IU and RDA Data Share Fellow class of 2015, and Robert Quick, OAB member of RDA and manager of High Throughput Computing at IU, both served as instructors.

3.2.3. PRAGMA

D2I is actively involved in the Pacific Rim Application and Grid Middleware Assembly (PRAGMA) and continues to contribute to the assembly growth and success. It organized the first Int'l Clouds for Data Science (ICDS'15) http://d2i.indiana.edu/pragma/ associated with PRAGMA 30 in Depok Indonesia. ICDS'15, which surfaced the research that is being carried out in PRAGMA, selected 8 papers for publication as part of a special issue in Concurrency and Computation: Practice and Experience.

PRAGMA 30 workshop in the Philippines showcased early work from the PRAGMA Rice project funded by an RDA/US Adoption award discussed above.

3.2.4. Midwest Big Data Hub

D2I organized IU's response to the successful NSF funded Midwest Big Data Hub (MBDH). MBDH is led out of NCSA with co-leads at IU, U of Michigan, Iowa State, and University of North Dakota. Plale leads the Data Science Ring; Bernice Pescosolido leads the Network Science Spoke. D2I is organizing a Fall '16 workshop funded in part from an award from the CRA CCC called "Data Quality in an Era of Big Data". The workshop will be held at Indiana University and take advantage of the facilities at CIB.

D2I is also growing the interest in Big Data on the IU campus, beginning with a summer'16 visit to SSRC, http://ssrc.indiana.edu/home/, to discuss opportunities with Director Emily Meanwell.

3.2.5. Smallholder Agriculture

In 2016 the D2I team made a considerable technical contribution to the collaborative project that focuses on examining adaptation of smallholder farmers in Kenya and Zambia to climate change. The project utilizes innovative data collection techniques, including the use of phones and its texting services and special devices that collect environmental data (PulsePods). The software developed by the D2I team members continuously ingests data from the phone data collection cloud service TextIt and stores it in a database for further analysis. A specially developed dashboard (web interface) allows to monitor the quality of ingest, quickly visualize results, and download selected data for further integration and analysis.

3.3. HathiTrust Research Center

The HathiTrust Digital Library is a partnership between academic and research institutions that offers a digitized collection of millions of titles. The HathiTrust Research Center (HTRC) is the research arm of HathiTrust and facilitates usage of the HathiTrust Digital Library by enabling computation analyses of the digitized library collection.

- HathiTrust Research Center is a partnership between IU Libraries, PTI, and School of Informatics and Computing at IU, and the University of Illinois, Urbana-Champaign (UIUC) Libraries and Graduate School of Library and Information Science."[2] HTRC's current focus is providing tools for analysis of large-scale texts – hundreds of thousands of documents comprising millions of pages of text. HTRC provides a portal from which researchers can operate text analysis applications using the some 3 million HathiTrust volumes that are in the public domain. HTRC also offers an API into the HTRC tools, so researchers can create their own text analysis tools.
- HathiTrust Research Center benefits from D2I cyberinfrastructure innovations that deliver optimal access and use of the HathiTrust corpus. The sheer size of the corpus demands innovative thinking about architecture and optimization at all levels of the software infrastructure from hardware to applications. D2I's research in this area focuses on reduced reads, intelligent caching, delivering maximum cycles at minimal costs, and providing secured environments for copyrighted materials. HTRC deals with 4.6 billion pages of printed material – so computational effectiveness and efficiency are critically important.

During this year, through persistence and support from IU and D2I, HTRC overcame the monumental technical, legal, and architectural barriers to ingesting and securely provisioning access to the entire HathiTrust (HT) Digital Library Corpus (over 14 million texts and 5 billion pages) for computational analysis. The effort involved cooperation and support by Research Technologies/PTI, IU office of legal counsel (OVPGC), IU Office of Research Administration (ORA), the IU Libraries, and the School of Informatics and Computing (SoIC).

Prior to the ingestion of full content, HTRC passed extensive internal security review and improvement process, codified in the HTRC document "Security: Measures, Practices, and Policies" (S:MPP) that was formally accepted by HathiTrust. The S:MPP implements the requirements set forth in the Collaboration Agreement executed by University of Michigan, IU, and University of Illinois at Urbana Champaign on behalf of HathiTrust and HTRC.

Full content from HathiTrust has been ingested into IU's Data Capacitor 2 in April 2016. The D2I team has successfully implemented the new version of Data Capsule – a secure virtual environment that enables computation while protecting the copyrighted content.

HTRC was also very successful this year in engaging researchers from various disciplines. It now has over a thousand of registered users from more than 200 institutions. Another clear evidence of successful engagement is that Andrew J. Mellon Foundation has made a several million-dollar commitment to HTRC, funding the integration of worksets – a framework that allows to package, process and share the results of non-consumptive research – with Data Capsule developed at HTRC.

In 2015 the HTRC funded four Advanced Collaborative Support Projects, which resulted in a number of findings and lessons learned. In 2016 the HTRC announced funding of four more projects, focusing on supporting library-research collaborations.

3.4. Digital Science Center

The Digital Science Center (DSC) investigates new programming models for parallel multicore and grid/cloud computing; develops Web services, portal, and gateway technology used to facilitate scientific computing and collaboration; and creates new methods for studying, modeling, and simulating complex networks and systems in nature, technology, and society. The DSC is led by Distinguished Professor Geoffrey Fox.

[2] HathiTrust, "What is HathiTrust?"

- Fox and DSC led the NSF-funded International Summer School on Data Science for Scattering Reactions, a graduate workshop dedicated to theory and phenomenology (data science) of scattering theory and its application to data analysis of modern experiments involving strong interactions physics. The 2-week graduate workshop took place at the campus of Indiana University with speakers from USA, UK, Finland, Israel, and Italy. See the workshop website: http://www.indiana.edu/~ssrt/.
- Fox and DSC continue to lead the $5M NSF-funded project to develop middleware and high performance analytics libraries for scalable data science. The project will produce two types of key building blocks: Middleware for Data-Intensive Analytics and Science (MIDAS) and the Scalable Parallel Interoperable Data Analytics Library (SPIDAL). MIDAS and SPIDAL are motivated and tested by the applications in 7 different communities: HPC Biomolecular Simulations, Network Science and Computational Social Science, Computational Epidemiology, Computer Vision, Geospatial, Remote Sensing and Pathology Informatics.
- Fox and DSC continue to play a major role in the $12M NSF-funded Comet supercomputer, the world's first virtualized high performance computing cluster. Comet is led by the San Diego Supercomputing Center, with IU and the Digital Science Center as key partners. IU Distinguished Professor Geoffrey Fox, an expert in the use of virtual systems in supercomputing, is leading key elements of the virtualization technology used in Comet.
- Fox and DSC continue to lead the $900,000 NSF-funded project to develop a Rapid Python Deep Learning Infrastructure, an artificial intelligence approach to solving big data problems.
- DSC is at present working intensively with the IU Precision Health Initiative Grand Challenge project to provide the data analyses software systems required for this initiative to be successful in improving human health. See https://www.dsc.soic.indiana.edu.

3.5. National Center for Genome Analysis Support (NCGAS)

The mission of the National Center for Genome Analysis Support is to enable the biological research community of the US to analyze, understand, and make use of the vast amount of genomic information now available. NCGAS focuses particularly on transcriptome- and genome-level assembly, phylogenetics, metagenomics/transcriptomics, and community genomics.

Major accomplishments and outcomes for FY2016 are:

- **NCGAS receives NSF renewal award.** The NSF Advances in Biologic Informatics (ABI) program has awarded NCGAS a three-year renewal to continue supporting genomics researchers, who do NSF-funded research or research compatible with NSF funding, nationwide.
- **Consulting**. NCGAS in the current year completed a total of 741 short-term consulting engagements (those taking less than 4 hours of staff time to resolve) and 235 long-term consulting engagements (taking more than 4 hours of staff time to resolve). Many the long-term consultations were research collaborations that last months or even years. In such collaborations NCGAS staff became partners and play a critical role in scientific discoveries accomplished by scientists receiving NCGAS help.
- **Education, outreach, and training.** NCGAS completed tutorials and training and outreach activities attended by hundreds of attendees.
- **Exploring Jetstream applications to bioinformatics**. As part of the Jetstream acceptance tests NCGAS helped researchers at the University of Arkansas Fayetteville to establish, provision, and use Jetstream VMs, to complete analysis of both northwest endangered river fish species, and the distribution of rattlesnake species, using ddRAD. This was done in partnership with Jeff Pummell

at the Strategic Initiatives and User Services, Arkansas High Performance Computing Center. We expect there will be future opportunities to aid researchers, using Jetstream. In another capacity, NCGAS is working with groups funded by Information Technologies in Cancer Research (ITCR) to use Jetstream in their work.

- **Strengthened partnership with Pittsburg Supercomputing Center (PSC).** During the initial ABI grant, PSC was an unfunded partner, and regularly exchanged information on Galaxy implementation, software installation, and the GenePattern suite. Now, beyond general coordination of software suites (and computational resources), we are beginning to build a center for metagenomic analysis centered at PSC. PI Doak attended a NEON metagenomics advisory meeting two years ago, and hopes to offer that project bioinformatics support—partnership with PSC will allow us to pursue this.
- **NCGAS and Trinity reach widely across the US and internationally.** NCGAS and Trinity now support 233 institutions in the US (see Figure 1), and 545 institutions in 49 countries (see Figure 2).

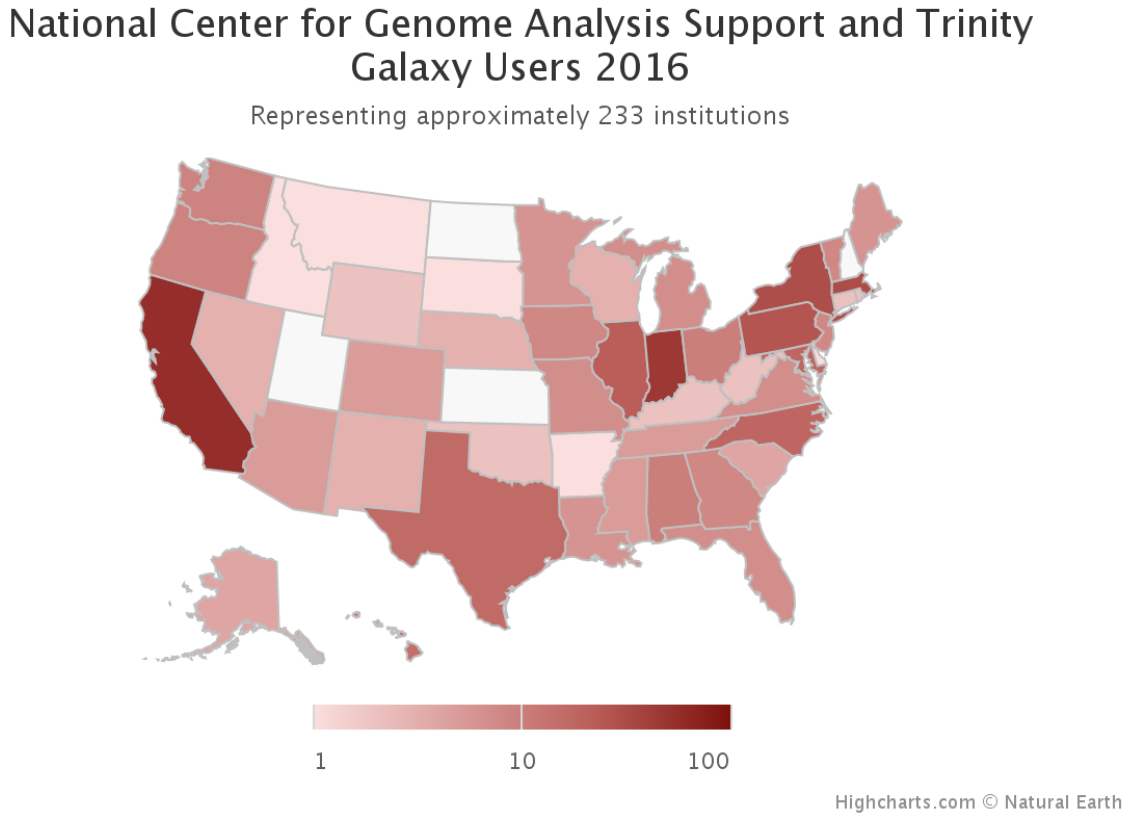


Figure 1. NCGAS and Trinity users in the United States.

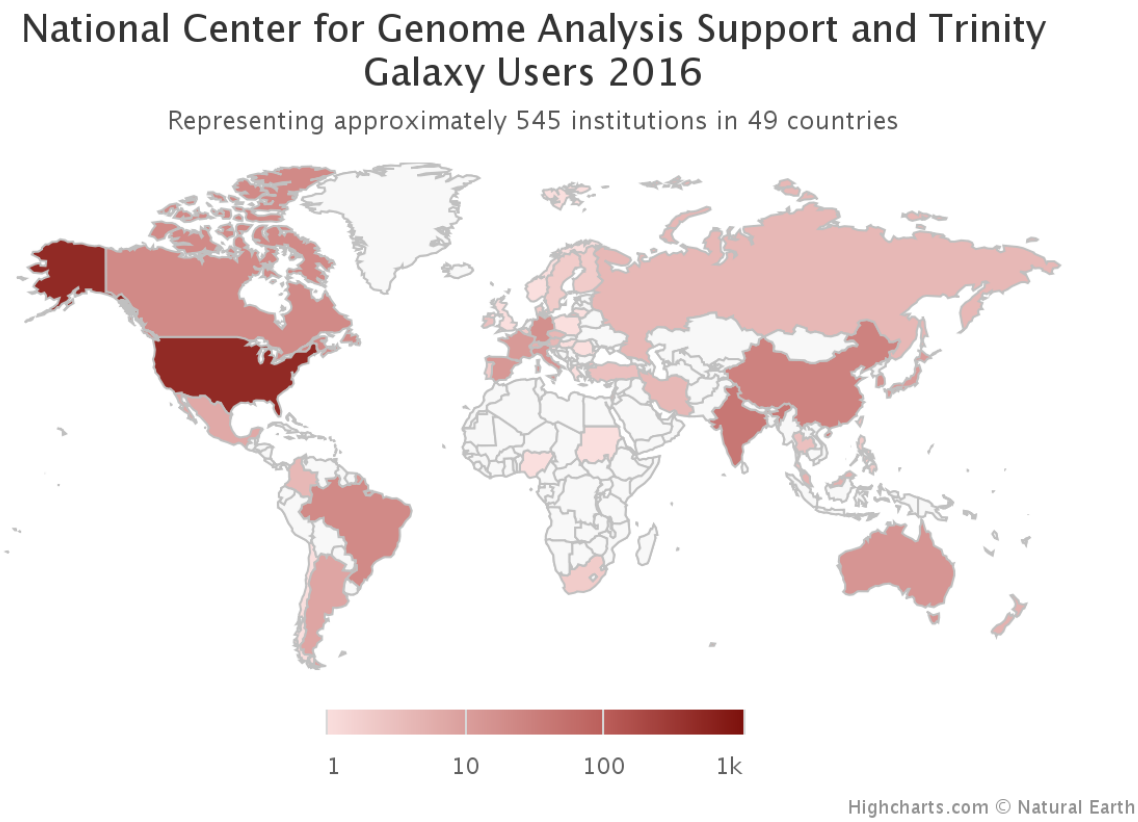
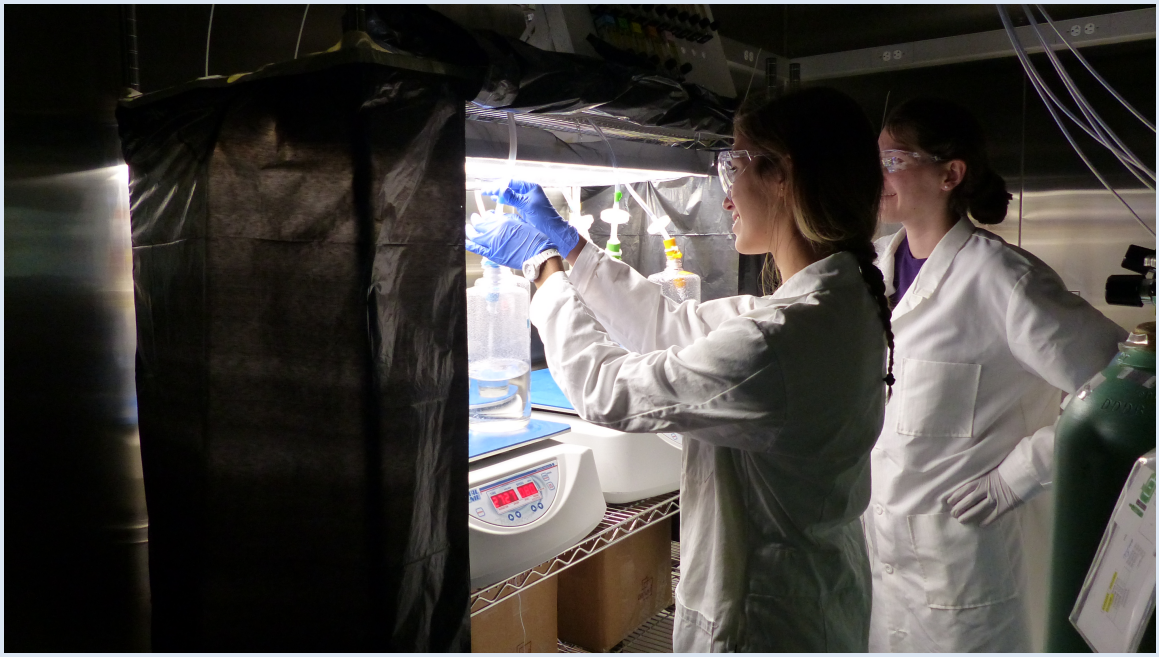


Figure 2. NCGAS and Trinity users worldwide.

Highlight: NCGAS enables improved understanding of ocean bacteria

Researchers Gwenn Hennon, Sheean Haley, J. Jeffrey Morris, Erik Zinser, and Sonya Dyhrman at Columbia University are investigating how microbial interactions modulate the effect of changing oceanic conditions on primary production (how plants make their own food) using IU's National Center for Genome Analysis Support (NCGAS) resources. Marine bacterial colonies respond to their surroundings, and things like the dissolved carbon dioxide (CO_2) and neighboring bacteria can have a large effect on a colony's success. Friendly neighbors can mean survival under conditions of double the normal amount of CO_2. By conducting analyses of gene expression on bacterial colonies, the researchers were better able to understand how a marine bacteria model system responds to elevated CO_2 levels.



Researchers at Columbia University preparing sample for analysis using NCGAS

These analyses allows for a deeper understanding of the effect of increasing CO_2 levels on oceanic life, as well as the complex relationships between species. The organisms studied here are at the base of the food chain and serve a vital role in ocean environments. These findings have significant implications for study of the impact of climate change.

3.6. Research Technologies

The mission of the Research Technologies division of UITS is to develop, deliver, and support advanced technology solutions that enable new possibilities in research, scholarly endeavors, and creative activity at Indiana University and beyond – and to complement this with education and technology translation activities to improve the quality of life of people in Indiana, the nation, and the world. As such, it is more of an engineering, delivery, and service organization than it is an innovation organization. However, pursuing excellent service for the IU community often involves innovating. A few key innovative projects led by Research Technologies during FY 2016 are described below.

- **Jetstream.** Jetstream was accepted as a production system by the National Science Foundation in FY 2016. Jetstream is a first-of-a-kind system; it is a configurable large-scale cloud computing resource that leverages both on-demand and persistent virtual machine technology to support a much wider array of software environments and services than current national cyberinfrastructure (National Science Foundation supported) resources can accommodate. As a fully configurable cloud resource, Jetstream bridges a major gap in the current XD ecosystem, which has machines targeted at large-scale, high performance computing, high-memory, large-data, high-throughput, and visualization resources. In particular, Jetstream:
    - Provides "self-serve" academic cloud services, enabling researchers or students to select a virtual machine (VM) image from a published library, or alternatively to create or customize their own virtual environment for discipline- or task-specific personalized research computing.
    - Hosts persistent VMs to provide services beyond the command-line interface for science gateways and other science services. Galaxy will be one of the initial science gateways supported.
    - Enables new modes of sharing computations, data, and reproducibility. Jetstream will support Digital Object Identifier (DOI)-based publication and sharing of VMs via Indiana University's persistent digital repository, IUScholarWorks, as well as supporting all Globus services, including data transfer/sharing with Globus Connect, and identity federation through Globus Nexus.
    - Expands access to the NSF XSEDE ecosystem by making virtual desktop services accessible from institutions with limited resources, such as small schools, schools in Experimental Program to Stimulate Competitive Research (EPSCoR) states, and Minority Serving Institutions. For example, VMs will enable use of Linux desktops from tablets.
- **Science Gateway Group provides science-enabling platform as a service and open source software.** SGG group manager Marlon Pierce serves as PI of the NSF-funded Science Gateway Platform as a service (SciGaP.org) project, which provides scalable, hosted services for science gateways developed by collaborators across the country. These services are based on the Apache Airavata software system. Apache Airavata is a top-level Apache Software Foundation project, and SGG team member Suresh Marru serves as Apache Airavata's official board representative to the Foundation. During the reporting period, SciGaP services enabled over 40 scientific publications by providing access to 13 different advanced computing resources around the world. Since 2014, the SGG has published over 20 peer reviewed papers and conference proceedings about science gateways and open source cyberinfrastructure.
- **Programmable Immersive Peripheral Environment Systems (PIPES).** (Programmable Immersive Peripheral Environment System) is a tool that extends commonly used virtual reality systems in support cyber-physical applications. PIPES allows a developer to programmatically control high voltage sockets. Students and staff have used PIPES in virtual environments to simulate wind, heat, and scent conditions thereby improving the immersive quality and increasing their suspension of disbelief. PIPES was developed by Chauncey Frend from the Advanced Visualization Lab (AVL) and was awarded the "Best Research Demo Award" by the IEEE Computer Society (2016 IEEE Virtual Reality conference in Greenville, SC). This is a major award in the area of advanced visualizations.

4. PTI Goal: Create capabilities with which researchers at IU (and beyond) associate and collaborate

4.1. PTI Sub-goal: Be the 'partner of choice' within IU and the nation for creating and implementing cyberinfrastructure facilities

IU's particular contribution to the concept of cyberinfrastructure is the role of people in enabling technology to support research, scholarship, and artistic creation. PTI brings together people across many organizational boundaries within IU, in order to serve the university and its constituencies.

This PTI sub-goal is tightly tied to Bicentennial Strategic Plan Goals Two, Three, and Five:

- Bicentennial Priority Two: A Community of Scholars
- Bicentennial Priority Three: Catalyzing Research
- Bicentennial Priority Five: A Global University

Detailed explanations of PTI contributions to these Bicentennial Priorities within the context of the overall goal of PTI being the "partner of choice" are described below.

4.1.1. Bicentennial Priority Two: A Community of Scholars

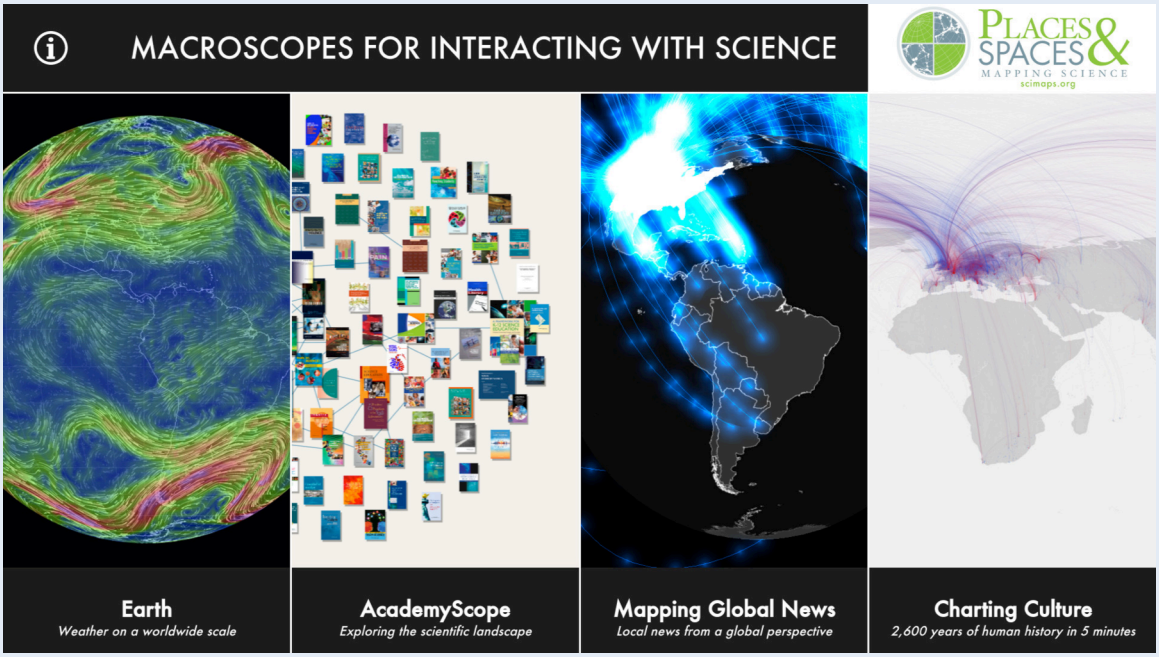
IU brings in computationally savvy junior faculty members and retains and nurtures them to full professorship here, and recruits such faculty to come to IU from other institutions. As or perhaps more importantly, IU attracts and retains faculty members who may not be computational science experts – supporting their use of advanced CI tools for research, scholarship, and artistic activity.

PTI, particularly the cyberinfrastructure facilities delivered and supported by the Research Technologies Division of UITS, is a highlight for Indiana University in recruiting and retaining top faculty talent. The role of PTI cyberinfrastructure is featured in a faculty recruitment video, "Supporting the building blocks of discovery: Indiana University's Advanced Cyberinfrastructure[3]."

[3] https://www.youtube.com/watch?v=cnlX6uhJVqI

Highlight: Macroscopes for interacting with science

The Advanced Visualization Lab (AVL) collaborated with the Cyberinfrastructure for Network Science Center (CNS) to present interactive data visualizations in a multi-touch kiosk. Interactive data visualizations, or macroscopes, have great potential as tools for exploring, understanding, and communicating science. The visualizations were submitted to the CNS by experts around the world as examples of "macroscopes" that help the user focus on patterns in data that are too large or complex to see unaided. The macroscope kiosk debuted at the David J. Spencer CDC Museum in Atlanta, Georgia on February 4, 2016 as the 11th iteration of the traveling exhibit, Places & Spaces: Mapping Science.



Examples of spaces and places using macroscopes

The macroscope kiosk was based on the design of the AVL's IQ-Table v2, a low-cost multi-touch table built with commercial-off-the-shelf hardware and open-source software. They empower individuals to interact with data and to make their very own maps. In the first 10 years of Places & Spaces, the exhibit was shown at 290 venues around the world and generated nearly 3.5 million website visits. The 11th iteration will spend five months at the CDC Museum in Atlanta, which hosts about 90,000 visitors each year.

Highlight: Angel Mounds interactive exhibit created by AVL and Glenn A. Black Laboratory

A collaboration between Advanced Visualization Lab (AVL) and Glenn A. Black Laboratory of Archaeology produced an interactive exhibit highlighting the Angel Mounds State Historic Site, located in Evansville, Ind. This exhibit is a high-resolution, multi-touch, tabletop display showing a map of the Angel Mounds sites and sheds new light on this prehistoric Native American site.

The creation of this new exhibit coincided with the Glenn Black Lab's 50th anniversary celebration. It will remain in the lab to educate visitors about the Glenn Black Lab, the Angel Mounds site, and the Mississippian culture. Due to the portability of the AVL's IQ-Table, the exhibit may travel to the Angel Mounds site in the future.

Visitors to the Glenn Black Laboratory of Archaeology at Indiana University interact with the Angel Mounds State Historic Site exhibit.

Metrics identified in IU Bicentennial Plan related to this Priority include:

- Number and subjects of continuing intercampus networks and number of faculty members involved. This metric is included in Table 4-D.
- Number of academic conferences held at IU that bring scholars here from other institutions. A list of academic conferences held at IU and organized by PTI is in
- Table 4-A.

Table 4-A. Academic conferences and workshops held at IU that bring scholars here from other institutions

Conference	Topic	IU attendees (faculty)	IU attendees (non-faculty)	IU attendees (total)	Faculty attendees from outside IU
Bioinformatics Clinic	Current methods in genome assembly	2	6	8	2
Intel Xeon Phi Workshop	?	5	43	48	2
BROAD/Trinity Kick-off Meeting	Genomics methods	1	4	5	1
Galaxy Community Conference Hackathon	Genomics computer hack	0	4	4	10
Galaxy Community Conference Training	Genomics analysis training	14	30	44	50
Galaxy Community Conference	Genomics research	14	34	58	50
Generic Model Organism Database (GMOD) project meeting	Genomics databases	5	1	6	18
Science Gateways Community Institute Kick-off	Science gateways	0	2	2	7
TOTAL		41	124	175	140

This year, PTI also aided in the management of conferences and workshops held at other locations, described in the table below.

Table 4-B. Academic conferences and workshops organized by IU and other collaborators, held at locations other than IU

Conference	Topic	IU attendees (total)	Faculty attendees from outside IU	Non-IU attendees (total)
Campus Bridging: Reducing Obstacles on the Path to Big Answers (September 2015)	Campus Bridging Workshop (held in conjunction with IEEE Cluster 2015)	1	5	10
IEEE/ACM Supercomputing Conference 2015 (November 2015)	IU hosted a booth with interactive demos on a variety of projects currently underway in UITS Research Technologies.	50	1,250	1,300
Visualization Infrastructure and Systems Technology Workshop (November 2015)	Workshop discussing the latest advanced visualization systems at SC15	4	75	150
Open Science Grid All Hands Meetings (March 2016)	Distributed High Throughput Computing	10	60	60
Gateway Computing Environments (GCE) 2015 (September 2015)	The GCE series, co-founded by RT's Marlon Pierce in 2005, provides a venue for science gateway developers and providers to make presentations of peer-reviewed research work in science gateway software development and operational strategies.	3	NA	50

Conference	Topic	IU attendees (total)	Faculty attendees from outside IU	Non-IU attendees (total)
CUG 2016 (May 2016)	The Cray User Group (CUG) is an independent, international corporation of member organizations that own Cray Inc. computer systems. Founded in 1978, CUG was established to facilitate collaboration and information exchange in the high-performance computing (HPC) community. IU staff led the overall event organization as chair of the CUG board.	8	4	231
LUG 2016 (April 2016)	The Lustre User Group is a meeting currently sponsored by the OpenSFS non-profit which brings together users and developers of the open source Lustre file system. IU staff led the overall event organization chairing both the program and organizing committees.	2	10	198
HUF 2015 (September 2015)	High Performance Storage System (HPSS) Users Forum is a meeting hosted annually by an HPSS customer site to engage customers, HPSS developers and other relevant HPSS staff for current issues and future plans discussions. IU staff were part of the Steering Committee and Technical Program Committee and annually lead the Burning Issues session (for tracking progress on customer reported issues).	3	0	100
XSEDE15 (July 2015)	The XSEDE conference series is a venue for researchers using XSEDE supercomputing and advanced research infrastructure, XSEDE support and operations staff, XSEDE outreach and campus representatives, and other interested groups to present research work, participate in Birds of a Feather and panel sessions, and organize tutorials.	28	Cannot be determined	596
XSEDE Science Gateway Symposium Series (July 2015 – June 2016)	The XSEDE Science Gateway Symposium Series is an online presentation series that presents a wide range of topics of interest to science gateway and scientific workflow developers and users. Archived presentations are available from https://www.xsede.org/gateways-symposium.	5	Cannot be determined	200
TOTAL		103	1,450	3,248

4.1.2. Bicentennial Priority Three: Catalyzing Research

IU's cyberinfrastructure (CI) – and the entire fabric of PTI's organization – is based on reducing barriers to innovation, discovery, improved health and quality of life, and artistic expression. A significant portion of PTI's activities is related to what can be considered grand challenges. The formalization of this initiative within IU will enable the Research Technologies Division of UITS, UITS as a whole, and all of PTI to focus CI research, development, deployment, and support activities on enabling IU to have significant national and international impact on the grand challenge topics selected for university-wide attention and investment. Already in FY 2015 Vice President for Information Technology Brad Wheeler has created a blue-ribbon faculty taskforce to provide guidance for the future of IU's investments in and support of CI, mindful of the need to support the IU grant challenge program. This new 2015 CI Research Taskforce will create a strategic report that will assess progress made by UITS and PTI against a similar

strategic taskforce report from 2005[4] and make strategic recommendations intended to guide IU in its research CI investments from now until 2025.

Some activities related to Bicentennial Priority Three – particularly items related to investment in *"…the physical and IT infrastructure necessary for twenty-first century research and beyond, through new and renovated laboratory space and continuing investments in cyberinfrastructure, with priorities based upon prospects for research productivity"* – are described later in this report. In this section of the document, we focus on PTI activities related to being the partner of choice. These are also relevant to Bicentennial priority three and the metrics defined for it, including:

- Amount and diversity of sponsored research
- Publication rate and demonstrated impact of faculty according to field-appropriate measures (already described in Table 3-A)
- Establishment of Grand Challenge groups and associated research funding
- Number of multi-campus sponsored research collaborations

4.1.2.1. Contributions of PTI overall to IU sponsored research

Table 4-C shows grant awards to IU where a PTI Center was a formal partner (PTL & IUPUI 1999-2014) as well as sponsored program (research) awards to the PTI as an aggregate (in FY 2015 and FY 2016 as reported by the IU Office of Research Administration) In addition to this sort of partnering, the Research Technologies Division of UITS offers a wide variety of services that are used by researchers in many disciplines and many responsibility centers at IU.

Table 4-C. Awards to PTI (from inception of PTL in 1999 and in FY 2015 and FY 2016)

	Grant awards, total, to PTI
PTL & IUPUI 1999-2014	$81,489,841
PTI FY 2015	$10,053,716
PTI FY 2016	$16,546,321

4.1.2.2. Use of PTI / Research Technologies cyberinfrastructure systems and services and External Funding to IU

In FY 2016, IU's Office of Research Administration reported $614,059,533 in total awards (grants and contracts). Figure 3 depicts grants and awards to the entire university, subdividing the awards to IU according to use of IU cyberinfrastructure services by PI or Co-PI. Figure 4 depicts the same sort of grant and award information for the IU School of Medicine and other Clinical Affairs Schools (IU Schools of Dentistry, Health and Rehabilitation Sciences, Nursing, Optometry, Public Health – Bloomington, Richard M. Fairbanks School of Public Health – Indianapolis, Social Work).[5] Grant and contract receipts were subdivided into four categories:

- No usage: no one (PI or Co-PI) with an account registered on any CI system or service supported by Research Technologies Division of UITS

[4] Wheeler, B.C. (ed.). Final Report of the Indiana University Cyberinfrastructure Research Taskforce. Indiana University. http://hdl.handle.net/2022/469

[5] http://www.iu.edu/initiatives/clinical-affairs.shtml

- HPC only: at least one person (PI or Co-PI) with an account on IU's high performance computing (HPC) and cyberinfrastructure services (Big Red II, Karst, Mason, Scientific Data Archive, Research File system)
- BIO: at least one person (PI or Co-PI) who uses databases, data resources, or collaboration tools delivered and supported by Research Technologies Division of UITS
- HPC and BIO: PI and Co-PIs include users of HPC and data /collaboration tools

IU's cyberinfrastructure contributed significantly to IU's grant competitiveness. It is certainly not the case that each and every grant award to a faculty member using PTI and RT resources depended critically on those resources. On the other hand, it seems impossible to believe that these resources are unimportant when almost half of the university's newly awarded research dollars went to people who use these facilities. To get a sense of the importance of these facilities to the research finances of the university, we need to look a little deeper at grant funding processes themselves. Most grant awards consists of two parts: direct costs – the cost of actually doing the research proposed – and indirect costs – which are calculated as part of a formula that takes into account the cost of buildings, research instruments, and cyberinfrastructure. For Indiana University, the indirect rate for research grants is 56% – meaning for every dollar in direct research costs, facilities and administration monies are calculated at 56%. (For service contracts and grants, the indirect rate is 32%. And there are some grant awards that come without any indirect funding.) In FY 2015, funding for indirect costs was more than $150 million. This is much higher than the $10M or so that IU puts in to research cyberinfrastructure each year. We know from federal grant agency feedback that IU's cyberinfrastructure helps the competitiveness of IU researchers applying for grants. And the $10M that IU invests in research cyberinfrastructure a year is a small fraction of the funding that IU receives!

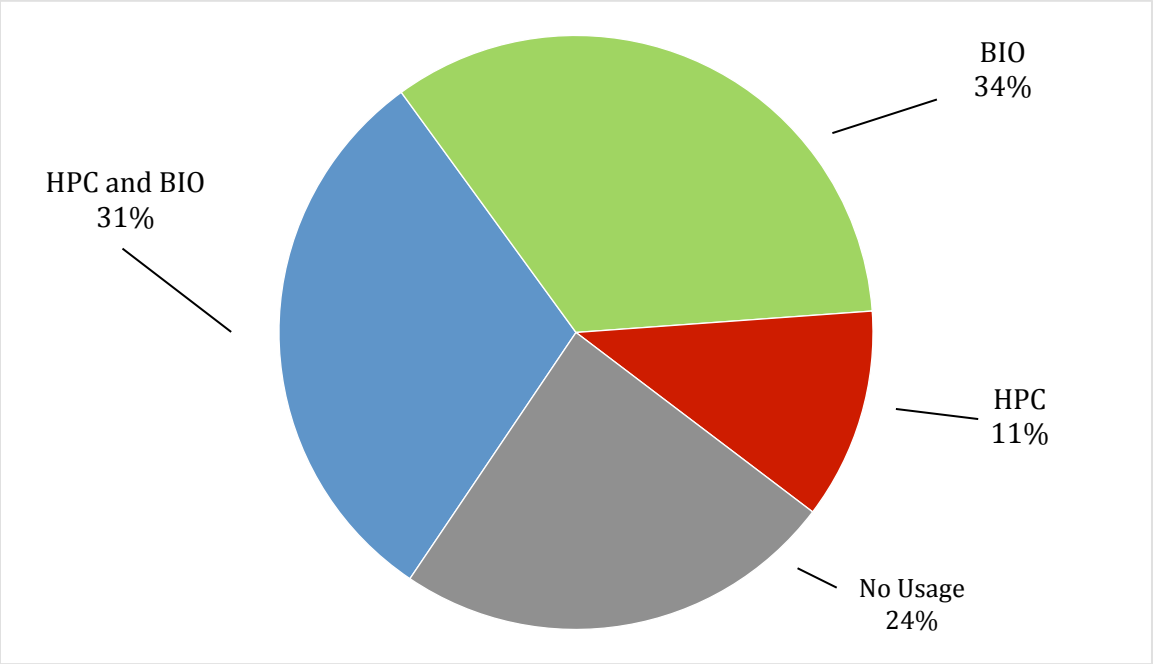


Figure 3. Grant funding to IU in FY 2016 associated with use of high performance computing systems (HPC), biomedical collaboration and data management services (BIO), or both.

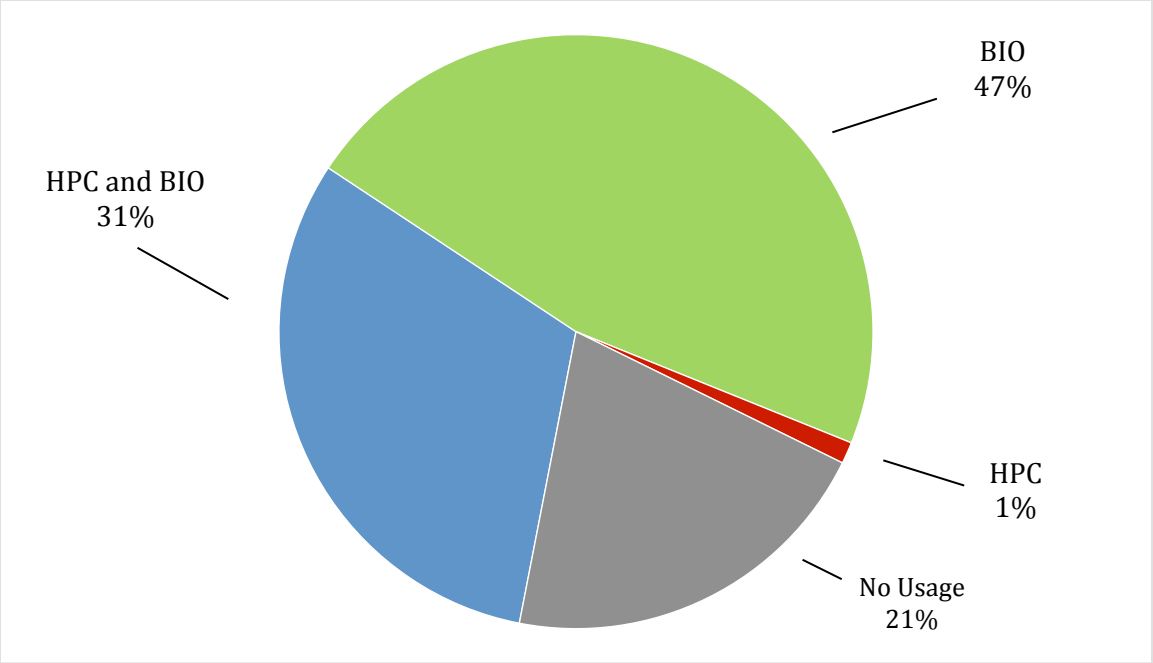


Figure 4. Grant funding to the IU School of Medicine and other Clinical Affairs schools associated with the use of high performance computing systems (HPC), biomedical and data management services (BIO), or both.

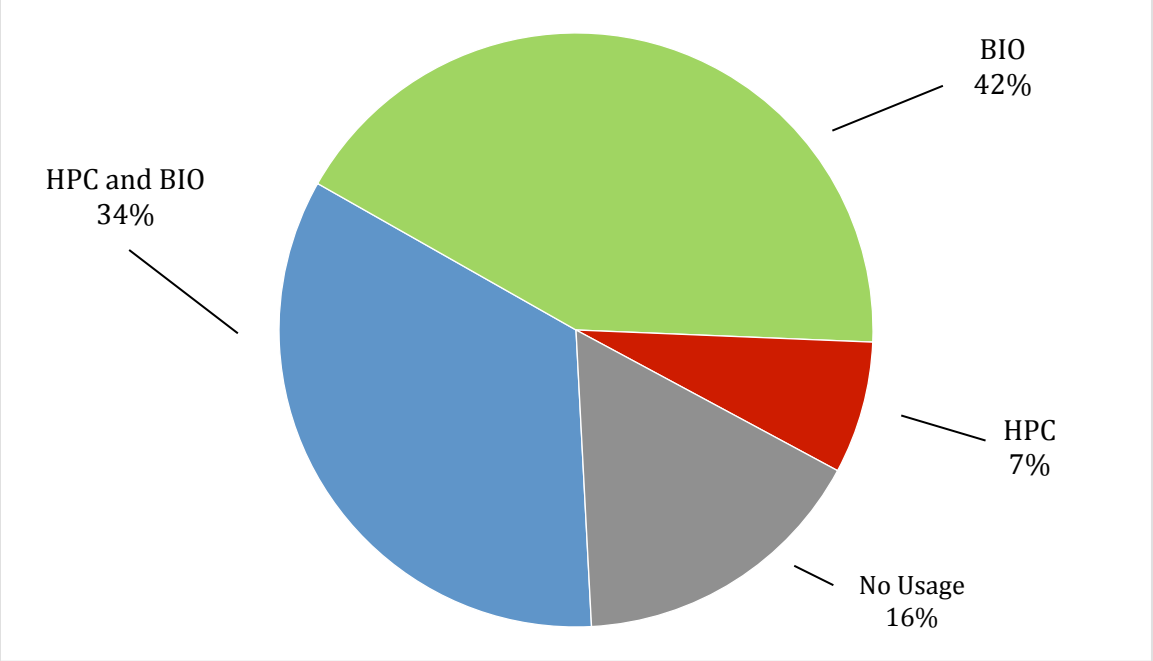


Figure 5. Grant funding from the National Institutes of Health associated with use of high performance computing systems (HPC), biomedical and data management services (BIO), or both.

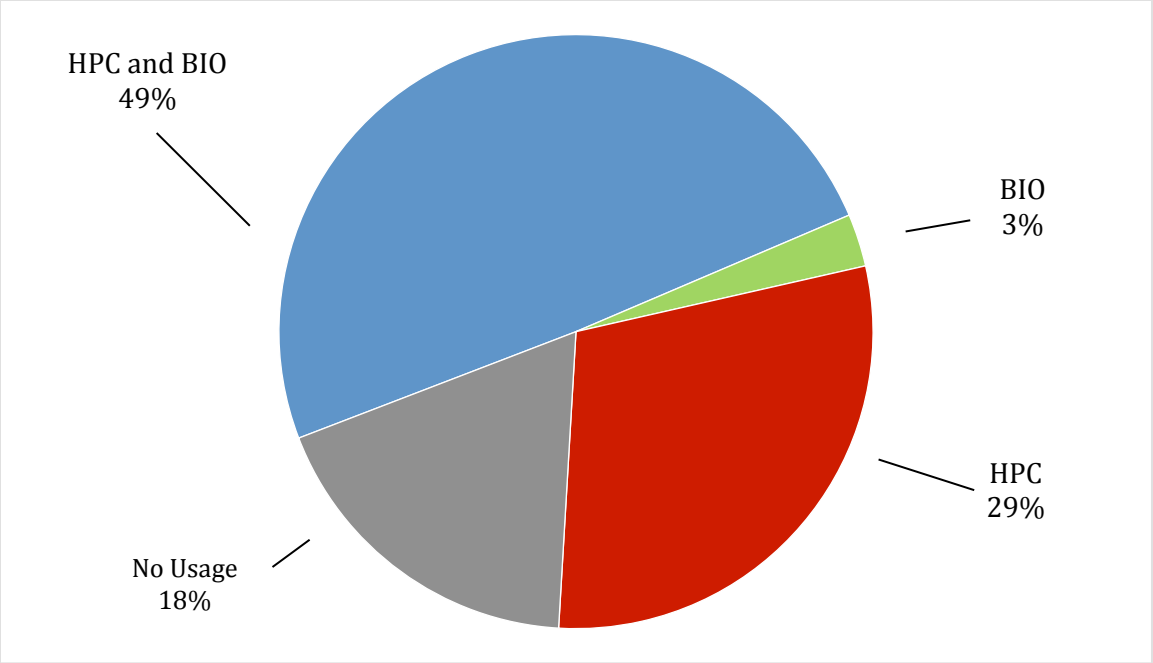


Figure 6. Grant funding from the National Science Foundation associated with the use of high performance computing systems (HPC), biomedical and data management services (BIO), or both.

4.1.2.3. Supporting research collaborations

The table below shows multi-campus sponsored research collaborations supported by PTI and the number and subjects of continuing intercampus networks.

Table 4-D. Number and subjects of multi-campus sponsored research networks led or supported by PTI

Name	Topic	Centers involved	PTI role		IU participation			Non-IU participation	
			Lead	Support	Faculty	Non-faculty	Campus	Faculty	Campus
CTSC	Cybersecurity for NSF science	CACR	X		0	8	1	1	4
SWAMP	Software assurance	CACR, RT		X	0	4	1	2	3
RDA	Informatics and computer science	D2I		X				> 4,000	Many international members
HTRC	Digital library	D2I	X		3	30	2	5	University of Michigan
SEAD	Scientific data preservation	D2I	X		1	20	1	20	UIUC, The Interuniversity Consortium for Political and Social Research
CTSI	New medical treatments	RT		X			IUPUI, IU	-	-

CIFASD/Imaging Core (AVL)	Fetal Alcohol Spectrum Disorder	RT		X	1	2	IUPUI	9	UCSD, SDSU, UCLA, Rivne Hospital, Emory University, Sanford Health, University of Minnesota, Khmelnitsky Hospital, University College London
NGVB	Gene therapy	RT		X			IUSM		Many U.S. Institutions
ADNI	Alzheimer's disease causes	RT		X			IUSM		59 acquisition sites;
OSG	Multi-disciplinary	RT		X	2	10	IUB, IUPUI	15	UW, UCSD, UNL, UIC
XSEDE	Multi-disciplinary	RT		X	6	6	IUB	620	UIUC, SDSC, PSC, TACC,
SciGaP	Biophysics, bioinformatics, computational chemistry, engineering, material science	RT	X		0	10	IUB		UTHSCSA, UCSD
Geo-Gateway	Geophysics	RT		X	0	10	IUB		NASA JPL, UC-Davis, UC-Irvine
NCGAS	Multi-disciplinary		X		0	0	0	0	
SPPEXA	Exascale Computing	RT		X			IUB	> 100	Many international members

DOSAR	Multi-disciplinary	RT		X		3	IUPUI, IUB	> 25	Bellarmine University, Iowa State University, University of Johannesburg, Langston University, LSU, Louisiana Tech University, University of Mississippi, University of Oklahoma, Universidade Estadual Paulista (UNESP) (SPRACE, GridUNESP), University of South Alabama, Sesquehanna University, University of Texas at Arlington
OSG	Multi-disciplinary	RT		X				> 1000	Many US locations

Highlight: RT supports research on polar ice sheets

Polar research supported by RT has resulted a better understanding of the interactions between the sea ice and land ice. The figure below shows an aerial photo of the Jakobshavn Glacier from the Operation Ice Bridge mission.

A roughly central portion of the massive and complex calving front of the Jakobshavn Glacier (photo by John Sonntag/NASA).

RT polar research involves support for field researchers creating significant amounts of remote sensing data. Our primary partnership in this realm is with University of Kansas' Center for Remote Sensing of Ice Sheets (CReSIS), in support of NASA's Operation Ice Bridge and other initiatives. Data collection efforts from these missions have secured:

- More than 300 flights total in the Operation Ice Bridge program
- Flights from April 18 – May 21, 2016 resulted in 29TB data collected

IU's activities in this realm offer considerable opportunities for contributing to the support of grand challenge research in climate change. The data captured through activities in support of Operation Ice Bridge provides a year-by-year record of Arctic and Antarctic change, and IU's technical expertise can support multiple partnerships performing field research in polar areas.

4.1.3. *Bicentennial Priority Five: A Global University*

Particular action items related to Bicentennial Priority Five include:

- IU will focus its international engagement efforts on its 32 priority countries and their leading institutions, establishing partnership agreements with leading or complementary institutions in all of these countries.
- IU will continue and enhance support for international students and for exchange programs that bring short-term faculty visitors to the university.

***Highlight:** IUanyWare enables international IU courses*

The IUanyWare team delivered tools and virtual desktops for an online course being taught in Budapest, Hungary in June 2016. This is the second summer that this program used IUanyWare technology to deliver the same course that Bloomington Kelly School of Business local students have privilege to access. The course in Budapest is now being delivered to 150 students via IUanyWare.

Table 4-E. PTI participation in international research and technical collaborations

International organization	Focus	Leaders from IU	PTI Centers Involved
Apache Software Foundation	Leadership of Apache Airavata, support for Apache incubating projects, student mentoring through Google Summer of code	Marru, Pierce	RT
Cray User Group	President, collaboration and information exchange in the high-performance computing (HPC) community	Hancock	RT
DOE Computer Graphics Forum	High-end visualization for HPC and simulation for National Labs	Sherman	RT
EAGLE	Program Committee	Gniady	RT
European Life-science Infrastructure for Biological Information (ELIXIR) – Computing Competence Center	Distributed Computing Infrastructure for Life Science Data Analysis	Quick	RT
European Grid Infrastructure – External Advisory Committee	Distributed Computing	Quick	RT
HPSS User Forum	Steering Committee Member, Technical Program Committee Member	Kallback-Rose	RT
InCommon Steering Committee	Advisor for Research	Welch	CACR
Internet Civil Engineering Institute	Director	Sons	CACR
Internet2 HPRC-PAG	Steering Committee Member	Stewart	RT
Linux Clusters Institute	Steering Committee Co-Chair	Knepper	RT
Lustre User Group	Program Chair, Promoting Lustre parallel file system and other open source file system technologies.	Simms	RT
Memoranda of Friendship with Technische Universität (Dresden, Germany)	Information services and high performance computing	Henschel	RT

International organization	Focus	Leaders from IU	PTI Centers Involved
Molecule to Brain (MoBrain) – Computing Competence Center	Bridging microscopic and macroscopic scale e-Science Tools	Quick	RT
National Oceanic and Atmospheric Administration Science on a Sphere Users Network	Research and educational institutions focused on application and advancement of Science on a Sphere technologies	Eller, Boyles	RT
Open Science Grid Council	High Energy Physics Research Consortium	Quick	RT
Open Science Grid Executive Board	Distributed High Throughput Computing	Quick	RT
Open SFS	Board Member (3 terms), nonprofit organization dedicated to the promotion and development of the Lustre file system	Simms	RT
SCxy 2006-2016	Executive Committee Member	Link	RT
Research Data Alliance	Data management and preservation	Plale, Quick	D2I, RT
SCxy Steering	Steering member 2012-2019, International conference for high performance computing, networking, storage and analysis	Link	RT
SPEC Consortium	Area Director, HPG, Standards body for performance benchmarks	Henschel	RT
SPXXL Leadership	Current Board Treasurer and prior Board Vice-President, Large-scale scientific/technical computing	Henschel, Hancock	RT
SPEXXA Leadership	Advisory Board Member	Stewart	RT

4.2. PTI Sub-goal: Enable the translation of software innovations to practical use

Highlights of PTI activities in translation of software innovations to practical use include the following:

- **Science Gateway Group supports NASA earthquake research.** For the last decade, the SGG has worked closely with NASA earthquake researchers to develop science gateways and cyberinfrastructure. The geo-gateway.org science gateway resulting from this work provides interactive online access to NASA data products and simulation tools. Geo-gateway.org services and SGG support have recently been used in a study of the 2014 La Habra, California earthquake that forecasts a 99.9% likelihood of an magnitude 5 or greater earthquake in the greater Los Angeles area in the next three years. The study also increases understanding of the earthquake fault system and its geophysical mechanisms underlying the greater Los Angeles area: slower moving processes are causing significant damage to the aging underground infrastructure (such as water mains) in Southern California.
- **Science Gateway Group leads the XSEDE Science Gateway Program.** The Science Gateway Group (SGG) leads and contributes significant staff time to providing web- and desktop-based cyberinfrastructure that simplifies access to XSEDE resources. During the reporting period, SGG staff members have served as lead or supporting consultants on nine XSEDE gateway projects. SGG staff members also organized three XSEDE15 tutorials in July 2015 on the topics of science gateway building, science gateway usage, and scientific workflows. Forty-five people attended at least one of the workshops. The workflow login node used for one of the workshops is now a persistent XSEDE resource and hosts four community-provided and maintained workflow tools.

- **CACR and HTC participate in securing software through the Software Assurance Marketplace (SWAMP).** CACR and the High Throughput Computing Group continue to provide the Software Assurance Marketplace (SWAMP), a $20 million DHS-funded facility that allows software developers and users to more easily identify and fix security vulnerabilities in their software, reducing the risks with using that software.
- **Center for Trustworthy Scientific Cyberinfrastructure.** Led by CACR, the Center for Trustworthy Scientific Cyberinfrastructure consortium is funded by NSF to lead its science community in securing the computational infrastructure critical to today's trustworthy science. In this role, CACR works with software development projects (e.g., Pegasus, SciGaP, Globus, NTP Foundation) to produce more secure software by advising them on good software development practices and the best way to implement new features.
- **NCGAS continues expanding access to genomic analysis software by researchers.** This includes the NSF-funded core purpose of NCGAS: to give researchers easy access to bioinformatics software packages on HPC systems capable of effectively running them, and to provide these packages in a menu-driven interface with Galaxy. Beyond pure research, NCGAS collaborates on two NIH Information Technologies in Cancer Research grants to make genomics analysis tools available to cancer researchers; the Trinity and GenePattern projects are centered at the Broad Institute and UCSF respectively. NCGAS works to improve these tools, but emphasizes making the tools broadly available.
- **AVL extends and further enhances its Collection Viewer application.** The AVL Collection Viewer software is an application built for collaboratively viewing and interacting with media collection(s). Media can include photos, videos, and audio clips. Commonly used in conjunction with the AVL's IQ-Table or IQ-Tilt systems, the software features an XML configuration file that allows exhibit creators to quickly and easily tailor the experience. Select new collections for FY 2016 included a beautiful exhibit featuring IU's Center for Network Science Places and Spaces Mapping Science collection, as well as a collection of African clothing from IU's Mathers Museum of World Culture. The Mathers exhibit debuted and was extremely well-received at the 2014 African Studies Association Annual Meeting in Indianapolis.
- **Open Science Grid Operations hosts network metric datastore and visualization for Large Hadron Collider Computing.** The OSG Operations group implemented a high availability database to collect performance-focused Service Oriented Network monitoring ARchitecture (perfSONAR) metrics for the Worldwide LHC computing facilities. This includes a Cassandra DB system that allows distributed high performance service management. The data collected can be immediately visualized via hosted services to troubleshoot active network issues, or historical data can be used to do long-term analysis of worldwide computing networking.
- **Text analysis with RNotebooks for beginners.** The CyberDH group has adapted algorithms from Matt Jockers into RNotebooks with markup that takes novices through the basics of text analysis (top ten counts, dispersion), working up to corpus clusters (dendograms) and using LDA. The code for this work currently resides on IU GitHub.

All told, PTI currently distributes a total of 31 open source software packages that are maintained on an ongoing basis and available for use by the US and global research communities as open source software.

5. PTI Goal: Impact the economic health and quality of life in Indiana – creating new jobs, nurturing new businesses

The mission of the Indiana University Pervasive Technology Institute (PTI) is to improve the quality of life in the state of Indiana and the world through novel research and innovation and service delivery in the broad domain of information technology and informatics. Thus, the PTI's mission is tightly related to IU Bicentennial Strategic Plan Seven.

5.1.1. Bicentennial Strategic Plan Goal Seven: Building a Prosperous and Innovative Indiana

Metrics associated with Bicentennial Strategic Goal Seven include the number and value of patents, licensing agreements, partnerships, and start-ups supported by IU. Continuing priorities related to Strategic Goal Seven include supporting regional business development in life science, technology, and related fields and facilitating university-industry collaboration, identifying opportunities to work in areas such as cybersecurity with Indiana defense-related institutions such as the Naval Surface Warfare Center (Crane), and the Indiana National Guard. The PTI supports these goals in several ways. For example, as shown in Table 5-A, the PTI supports the development of patents; in Table 5-B are the Indiana-based start-up companies and in Table 5-C are the non-startups with which the PTI engages.

Table 5-A. Patents awarded to faculty and staff with a relationship to PTI

Description	Patent #	Relationship to PTI	Web address
Wireless Network Radiolocation Apparatuses, Systems and Methods	**US 20090054106 A1**	Steven Wallace and Danko Antolovic were PTL staff members at the time of the patent filing	http://www.google.com/patents/US20090054106
Compression system and method for accelerating sparse matrix computations	**WO 2007095516 A3**	Professor Andrew Lumsdaine and academic research associate Jeremiah Willcock were affiliated with PTI at the time of the patent filing; patent partly owned by IURTC	https://www.google.com/patents/WO2007095516A3?cl=en&dq=lumsdaine+patent&hl=en&sa=X&ved=0CB0Q6AEwAGoVChMIvuj15rr0yAIVyCoeCh2XLgGu
Total – 2 Patents awarded with some relationship to PTI			

Table 5-B. Indiana-based startup companies and organizations established with some sort of engagement or assistance from PTI from inception of PTL in 1999 that are still active as of the end of FY2016. (Life sciences companies are indicated with an *)

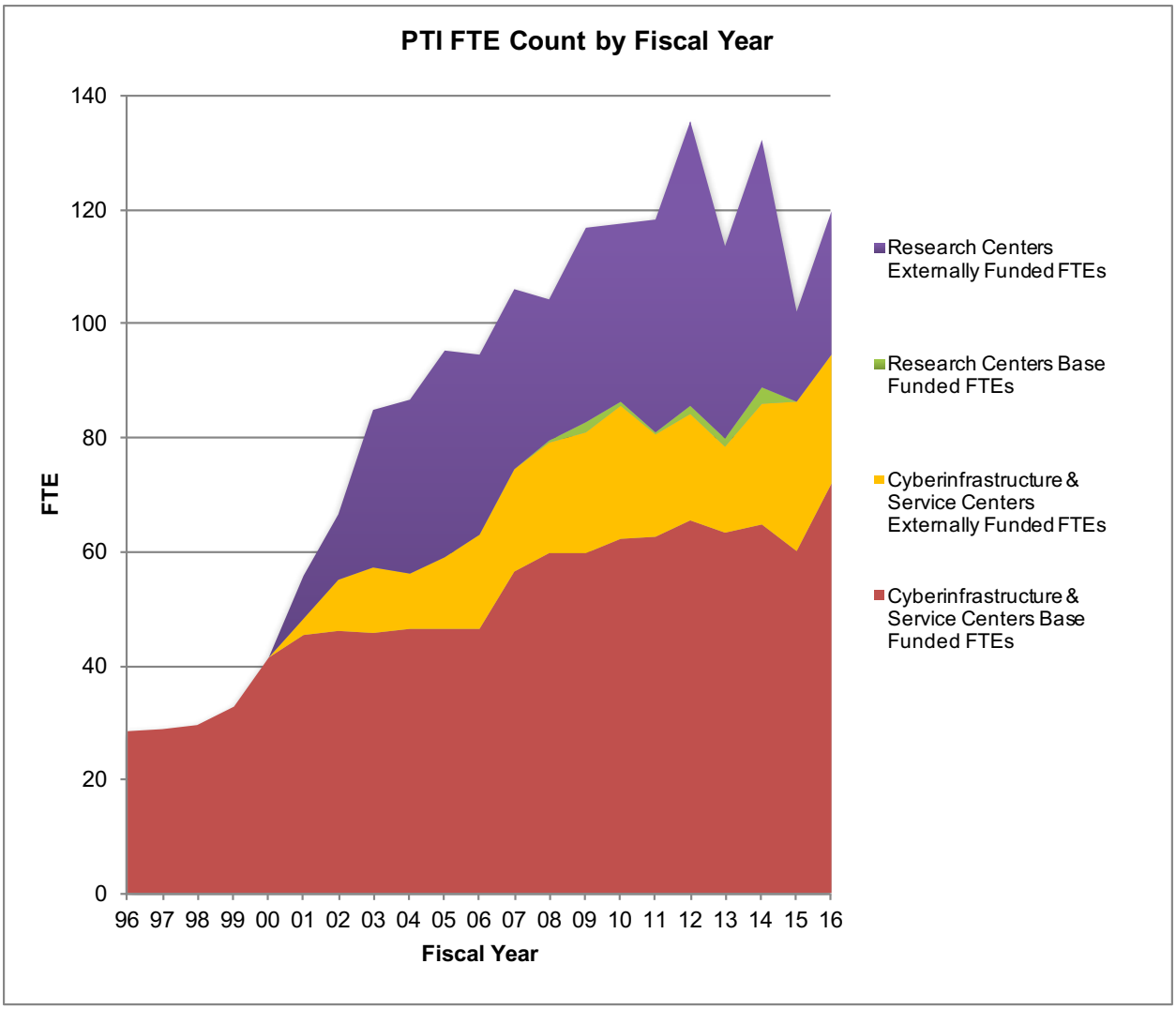
	Description	Relationship to PTI	Web address
Bloominglabs, a 501(c)3 not for profit which is Indiana's first makerspace.	Makerspace, robotics	Founded by a group of people including PTI staff, events annually sponsored by PTI.	http://bloomingtontech.com/project/bloominglabs/
Chalklabs	Data Analytics	Company founded by students of PTL Fellow Dr. Katy Borner	http://bloomingtontech.com/project/chalklabs/
Precise Path Robotics	Autonomous robotic lawnmowers	Started with IP (Intellectual Property) created by PTL staff	
Wisdom Tools	Educational games	Early leaders included faculty with PTL affiliations	http://www.wisdomtools.com
Total – 4 startups with some affiliation to PTI			

Table 5-C. Companies other than startups that operate within Indiana and which benefit from some sort of support or engagement with PTI from inception of PTL in 1999 that are still active as of the end of FY2016. (Life sciences companies are indicated with an *)

	Description	Relationship to PTI	Web address
Cigital	Cybersecurity consulting and analysis	Branch opened in Bloomington as a result of relationship between Cigital CTO and former PTL Science Director Dr. Dennis Gannon	http://bloomingtontech.com/project/cigital-inc/
Cummins Inc.	Porting computational fluid dynamics simulations to GPUs on Big Red II	Research Support	http://www.cummins.com/
Naut Inc.	Scalability testing of algorithms on Karst and Big Red II	Research Support	http://www.nautinc.com/
WS02	Open Source Enterprise Software Development; opened branch office in Bloomington	Strategic partnerships with PTI's D2I Director Beth Plale, DSC Director Geoffrey Fox, and RT members Suresh Marru and Marlon Pierce	http://wso2.com/
Total – 3 companies operating in Indiana with some relationship to PTI			

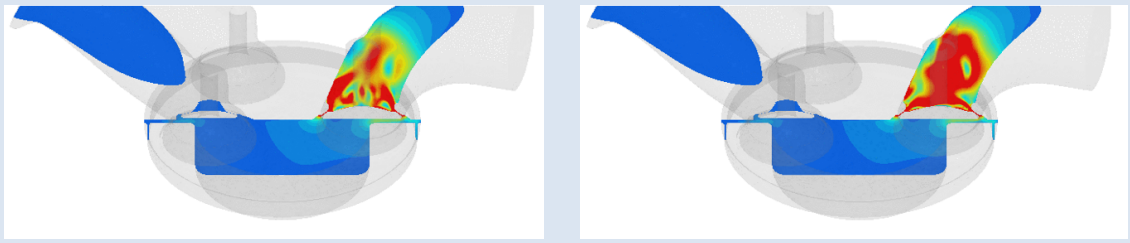
One of the most important ways in which PTI enhances the economic health and quality of life in Indiana is by being successful in the highly competitive process of winning federal grants and contracts, and creating new high quality jobs in Indiana as part of IU. Such jobs have an average salary greater than the overall average in Indiana, add to tax roles, and often bring highly qualified professionals from other states or nations to Indiana, where they very often settle down and stay their entire careers.

Figure 7. PTI employment over time, measured in Full Time Equivalents (FTEs), showing positions funded by external grants and contracts.


PTI FTE Count by Fiscal Year
140
120
100
80
60
40
20
0
FTE
96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16
Fiscal Year
Research Centers Externally Funded FTEs
Research Centers Base Funded FTEs
Cyberinfrastructure & Service Centers Externally Funded FTEs
Cyberinfrastructure & Service Centers Base Funded FTEs

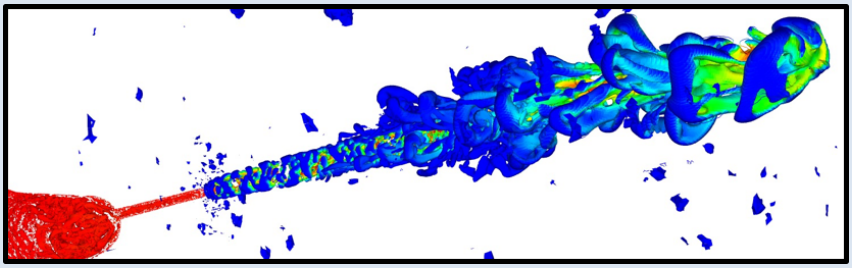
Highlight: SciAPT group and Cummins improve combustion simulations

IU's Science Applications and Performance Tuning group assisted project leads John Deur and Siddhartha Banerjee from Cummins, Inc. (Columbus, IN) by using graphic processing units (GPUs) to calculate reaction rates. Without this GPU-based approach, solving the combustion chemistry of complex fuels inside a combustion chamber is impractical. With the GPUs, it is now possible to improve the ignition predictions of new fuels and understand the formation of soot pre-cursors.

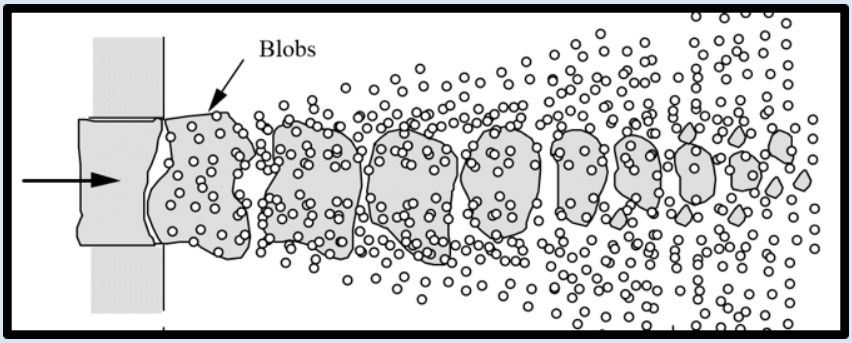


RANS (traditional computational fluid dynamics) vs. LES (HPC-based method)

As a result of this work, Cummins has considerably improved its designs and understanding of fluid behaviors.



Spray modeling and primary atomization of liquid jet (HPC-based modeling)



Spray modeling and primary atomization of liquid jet (traditional approach)

This work was made possible by an earlier partnership of Lawrence Livermore National Laboratory, Indiana University, and Cummins to adapt engine simulation software from Convergent Science (the top maker of engine simulation software) to GPUs.

6. PTI Goal: Support the development of a 21st century workforce within the State of Indiana

PTI supports the development of a 21st century workforce in a variety of ways. One of the primary ways is through education offered by faculty with a PTI affiliation. As stated earlier, course delivery activities result in academic credits at IU, and degree-granting programs are offered primarily through PTI-affiliated centers that are also subunits of the IU School of Informatics and Computing (D2I and DSC in particular). PTI provides a number of services and facilities that support the development of a 21st century workforce in Indiana by facilitating success and innovation by current IU students, and by spurring interest in STEM disciplines (Science, Technology, Engineering, and Mathematics) through K-12 outreach programs.

6.1.1. Bicentennial Priority One: A Commitment to Student Success

The PTI goal of supporting the development of a 21st century workforce is closely aligned with the IU Bicentennial Priority One. Current critical services offered to undergraduate and graduate students of IU include access to software (analytical, geographic information system, mathematical, and statistical) and access to and use of supercomputers. Graduate and undergraduate students have open access to Karst, IU's newest computing resource brought online in October 2014 (and prior to that a similar resource, Quarry). These resources are available without usage fees to support research, artistic, and scholarly endeavors related to projects in the classroom, the lab, or art and design studio. Similarly, IU's largest computing resource, Big Red II, and a specialized large-memory system, Mason, are both open to graduate students by default and undergraduate students with a faculty sponsor.

- **Access to and use of database systems.** The IU research database environment is described in Appendix 1. As with IU's supercomputing resources, graduate and undergraduate students have open access to the Research Database Complex (RDC) to host data in an Oracle or MySQL relational database without usage fees. Students can also leverage the RDC as a backend to web applications hosted within the environment or elsewhere.
- **Use of advanced storage systems.** IU's data storage systems are described in Appendix 1.
- **Access to the SDA.** Graduate students and sponsored undergraduate students have access to the Scholarly Data Archive, which provides large capacity, long-term storage in IU's tape-based archival system. The 50TB default quota is extended students, faculty, and staff alike.
- **Access to DC2 and DC-WAN.** IU students, faculty, and staff using IU's HPC resources by default receive access to Data Capacitor II (DC2) and Data Capacitor WAN (DC-WAN), high capacity, high bandwidth file systems for the short- to mid-term storage of research data. Both systems use the open source Lustre file system, present on over 60% of the world's top 100 supercomputer systems. DC-WAN permits a file system mount across long distance to enable geographically distributed workflows and new ways of collaborating.
- **24/7/365 use of the entire suite of advanced visualization technologies.** Available technologies include ultra-resolution IQ-Walls, interactive and collaborative IQ-Tables and IQ-Tilts, 3D scanning equipment, spherical displays, and a variety of interfaces and displays that support virtual and mixed reality. Of particular note, IU Bloomington has seen increased student use of IQ-Walls in public spaces (e.g., the main library). IUPUI has seen tremendous uptake from students because of their close partnerships with Informatics. Informatics students met and utilized the provided visualization technology nearly daily, and often on weekends and evenings. A loaner program for portable equipment (the most popular being the Oculus Rift Development Kits) affords additional opportunities for independent learning and exploration.

- **Use of the GitHub Enterprise Service.** IU is the first academic institution to provide the GitHub Enterprise Service for distributed source code control, software development, and collaboration. The git distributed version control system allows considerable flexibility and ease of collaboration in a networked environment. This service has seen considerable uptake across the university, not only from the School of Informatics and Computing, but also in numerous academic units as well as in administration. Currently github.iu.edu hosts 19,073 repositories owned by 6,656 users, supporting student projects and assignments, departmental web services, faculty collaborations, and enterprise software development.
- **Access to and use extensive suite of software.** There are basically five ways that students can use the Stat/Math software: 1) in STC labs, 2) on IUanyWare, 3) on central research systems, 4) purchasing a copy from the bookstore, and 5) purchasing a copy directly from Research Analytics. The numbers in Table 6-A only take into account a subset of the software titles (from central research systems) –as it currently stands, only usage of Maple, Mathematica, Matlab and SPSS on the central research systems (BR2, Karst and Mason) are included. Additionally data for SAS, Minitab, etc. could not be obtained as these software packages do not use a license server. the actual number of student users almost certainly multiple thousands.

Table 6-A. Student users of IU's advanced cyberinfrastructure)

Type of system	Undergraduate students		Graduate students		Total	
	FY 2015	FY 2016	FY 2015	FY 2016	FY 2015	FY 2016
Users of statistical and mathematical software	40*	26	186*	164	226*	190
IUanyWare	--	>55,000	--	>5,000	--	>60,000
Supercomputers & computational systems	36	23	558	331	605	354
Advanced storage systems	282	58	1,157	484	1,477	542
Database systems	11	3	42	27	53	30
Advanced visualization systems	144	168	73	70	217	238
RT GitHub code repository	1,640	1,638	1,154	1,189	2,794	2,827
Google Summer of Code participants	0	0	4	8	4	8

*Maple, Mathematica, Matlab and SPSS usage on the central research systems (BR2, Karst and Mason).

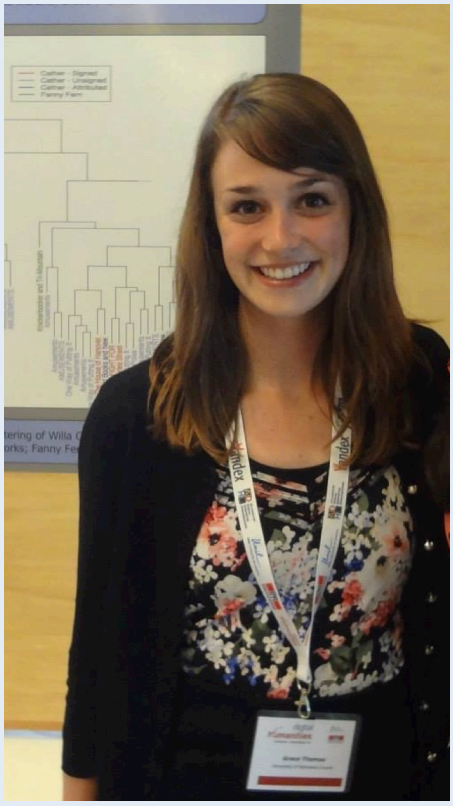
6.1.2. Employment, education, and practical experience for IU students

Table 6-B. Practical experiences for students in the PTI

Student	Program	Group	Project Description
Grace Thomas	MIS & MLS, Information and Library Science, SoIC	CyberDH	Created R scripts and notebooks to teach text analysis processes to students, faculty, and researchers
Gagandeep Sing	MS, Computer Science, SoIC	AVL	Developed lightweight, Web-based virtual reality, augmented reality, and interactive 3D tools.
Alex Doub	BS, Intelligent Systems Engineering, SoIC	AVL	Design and construct a prototype device that creates a low-friction walking surface that can be interfaced with a virtual reality display
Yan Zhou	Ph.D., Inquiry Methodology School of Education	Research Analytics	Student and faculty support for statistical analysis methods and software; UITS survey analytics; new ways to visualize and interact with the data using Tableau
Wenjuan Sang	Ph.D., Education Policy Studies	Research Analytics	Created a system that will aid Research Analytics in tracking software purchases and allow for more accurate reporting of sales across campuses
Scott McClary	BS, Computer Science, SoIC	SciAPT	Worked with an IU chemistry research group to improve their workflow and move it from a desktop to Karst. This improved their throughput by an order of magnitude.

Highlight: CyberDH supports student success

During the 2015-2016 school year, MLS/MIS candidate Grace Thomas worked with the Cyberinfrastructure for Digital Humanities team to expand the group's offerings on text analysis with R for humanists. Thomas came to the team having worked with Matthew Jockers at the University of Nebraska, the author of the highly influential book *Text Analysis with R for Students of Literature*.



Grace Thomas giving a presentation

The aim of CyberDH's materials are to ease humanists into programming with R by providing a three-tired approach to freely available algorithms: 1) a web-based Shiny app, 2) a fully annotated RNotebook explaining the how and why of each line of code, and 3) a lightly annotated RScript for adaptation by users. These materials are used in workshops given on the IU campuses and have been presented at two conferences, Keystone DH and XSEDE.

Thomas was instrumental to drafting initial versions of our algorithms & workflows and assisting at workshops. This work made her a highly desirable job candidate. Thomas is now a web archivist at the Library of Congress.

Congratulations Grace!

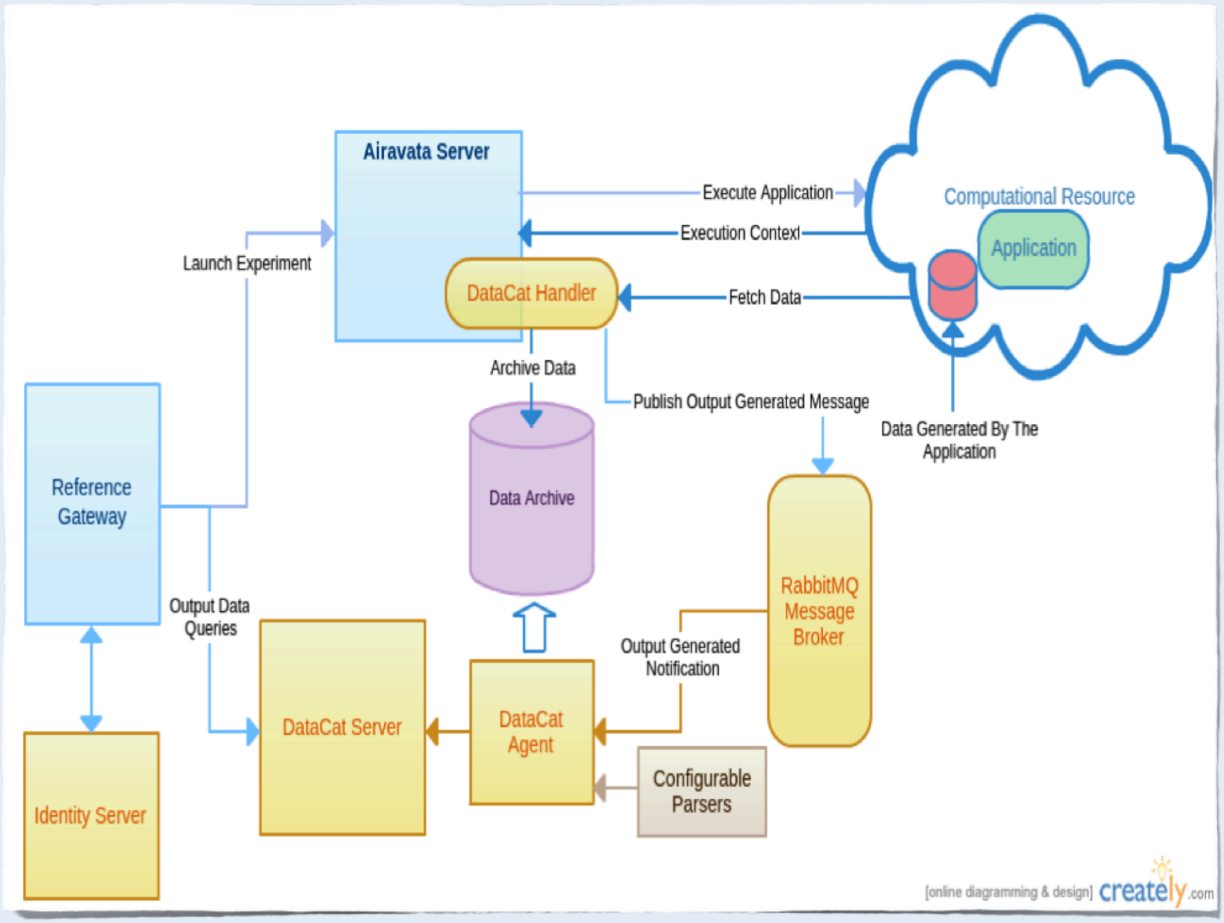
Overview of 2015 Student Projects

- Integrate SEAGrid GUI with Apache - Dimuthu Upeksha, University of Moratuwa, Sri Lanka
- GenApp Integration with Apache Airavata - Abhishek Kapoor, Priyanshu Patra, Indian Institute Of Technology (IIT) Kharagpur
- Integrating DataCat system with Apache Airavata - Supun Nakandala, University of Moratuwa, Sri Lanka
- Securing AIRAVATA API - Hasini Gunasinghe, Purdue University
- Evaluate Apache Airavata Metadata storage and explore alternative solutions - Douglas Chau, State University of New York, Binghamton
- Integrating Apache Mesos with Science Gateways via Apache Airavata - Pankaj Saha, State University of New York, Binghamton

- Benchmarking Resource Usage of Airavata's Applications - John Weachock, State University of New York, Binghamton

Highlight: Google Summer of Code - 2015

Google Summer of Code is a program that introduces students to open source software development. The students are paired with organizations and senior developers and architects to develop a project during the summer break. For several years, IU's Science Gateways group has led students from across the globe through the GSOC program. In 2015, there were eight students participating with the Science Gateways group. Most projects used Apache Airavata



Example of a project's architecture. All projects can be viewed here.

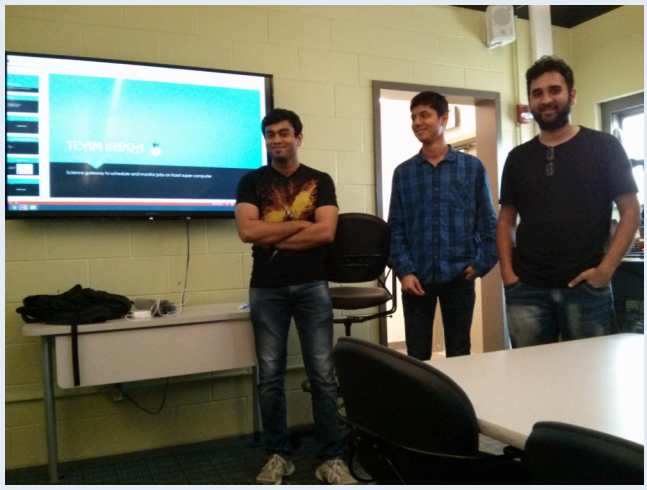
Highlight: AVL teaching students – enabling student success

The Advanced Visualization Lab (AVL) partnered with the IU School of Informatics and Computing (SOIC) to provide a unique course on visualization. The AVL's course focused on the basics of virtual reality. Students learned how to create their own virtual reality applications which could be experienced on displays ranging from Smartphone viewers, large tiled display systems and an Oculus Rift head-mounted-display. Students used the stereoscopic tiled wall at the Scholars' Commons in the Wells Library, an Oculus Rift, and "Google Cardboard" (provided by Google Inc.) for project development.

Students and an instructor using the stereoscopic tiled wall during a project demonstration.

Highlight: Science Gateways group offers unique course to IU students

The Science Gateways group partnered with the IU School of Informatics and Computing (SOIC) to provide a specialized course on science gateway architecture. In this course, the students formed development teams, and each team built a science gateway software as a service system from scratch. Students learned DevOps principles for deploying robust cloud services. Students were also introduced to the Apache Software Foundation's open community governance principles for open source software and will learn how to effectively interact with Apache Software Foundation projects in order to become committers and project management committee members.



Students using Jetstream in a team project demonstration

7. PTI Goal: Offer services that enhance enable new innovations and accelerate research by the IU research, scholarly, clinical, engineering, and artistic communities

The Research Technologies Division of UITS and the National Center for Genome Analysis Support (NCGAS) are the two centers of PTI referred to as Cyberinfrastructure and Service Centers. RT and NCGAS engage in activities particularly relevant to the following IU Bicentennial Priorities and Continuing Priorities:

- Bicentennial Priority Three (continuing priority of "The Centrality of Information"): Catalyzing Research
    - Pervasive deployment of IT. Ensure information technology is pervasively deployed at IU by leveraging and continuing the support of the university's long-standing and internationally recognized excellence in information technology services and infrastructure.
- Bicentennial Priority Six: Health Sciences Research and Education to Improve the State and Nation's Health.
- Bicentennial Priority Eight: Towards a Culture of Building and Making.

Relevant PTI activities – through the actions of the two cyberinfrastructure and service centers – have aided IU innovations in many ways, described in more detail below.

7.1.1. Bicentennial Priority Three / Continuing priority of "The Centrality of Information": Ensure that the Principles of Excellence are supported by outstanding information technology and information resources.

This continuing priority is further explained as: Pervasive deployment of IT. Ensure information technology is pervasively deployed at IU by leveraging and continuing the support of the university's long-standing and internationally recognized excellence in information technology services and infrastructure. There are a number of sub-items for this continuing priority identified in the IU Bicentennial Plan, each addressed below with two exceptions. Those exceptions are the following sub-items:

- Maintain IU's leadership role in network management and cybersecurity in Indiana, nationally, and internationally. This has already been discussed earlier under CACR activities.
- Implement Empowering People: IU's Strategic Plan for IT, with particular focus on systems for the use of institutional data, supporting student success, and facilitating academic processes.

7.1.1.1. Ensure that IU continues to provide an outstanding, flexible, and secure IT infrastructure for students, faculty, and staff

RT and subsequently CACR, with a transfer of responsibility late in FY 2016, bear primary responsibility for maintaining Health Insurance Portability and Accountability Act (HIPAA) alignment for UITS services. This includes periodic controls and documentation revision, staff training and workshops,

regular external audits, and coordination with Compliance, Security, Internal Audit, and the Data Stewards. It also encompasses establishing and maintaining institutional standards for securing electronic protected health information (ePHI), responding to incidents, and assisting UITS units as they put in place the appropriate controls and other requirements to sufficiently secure ePHI. This service has proven useful and needed by a number of units across UITS and has become part of the IT risk management landscape for IU.

PTI's activities in this have firmly established IU as a national leader in regulatory alignment of IT systems for HIPAA and FISMA (the Federal Information Systems Management Act). HIPAA alignment is important to UITS' research relationship with Clinical Affairs schools, researchers who work with ePHI on all campuses, and UITS' growing relationship to IU Health. The risk for non-compliance can be high: In 2013, a breach of the personal health information from 6,800 people at Columbia University and NY Presbyterian Hospital resulted in a fine of $4.8M.

During FY 2016, the following certificates were supported in achieving their HIPAA alignment: Box, CAS, Conveyant, IUanyWare, axiUm (on IUanyWare), New RFS (Geode), and WebServe/CHE.

7.1.1.2. Invest in cyberinfrastructure for education and research that emphasizes flexible and scalable high speed computation, massive data storage, and extensive high-speed network connectivity that enables education and research

When the UITS Research Technologies Division proposed the purchase of Big Red II, we made a commitment to offer training, information, and support services to increase the diversity of disciplines and sub-disciplines that make use of the system. In particular, we set a goal of having Big Red II used by at least 150 disciplines and sub-disciplines practiced at IU (out of 381 recognized categories). PTI met and exceeded this goal, as shown in **Table 7-A** below.

Table 7-A. Disciplines and sub-disciplines represented among users of Big Red II and Karst (rows and columns are not additive because the totals are unique disciplines and sub-disciplines, and there are several users)

System	IUPUI		IUB		Total	
	FY 2015	FY 2016	FY 2015	FY 2016	FY 2015	FY 2016
Big Red II	68	87	141	159	159	254
Karst	117	152	164	218	214	180
Total	136	168	201	238	243	279

Educational programs such as the "Supercomputing for Everyone" educational series provide very effective training on Big Red II. But where a researcher's needs or student's needs can be met by Karst, the fact that it is the less expensive of the two systems (on a per-unit-of computing basis) means use of Karst makes effective use of IU financial resources. In fact, as a result of faster clock rates, Karst completes many serial or single-node computational tasks in a shorter elapsed time than Big Red II. As shown in Figure 8, during FY 2016, CPU time devoted to parallel

jobs on 34 processors (17 nodes) or more on Big Red II varied around 60% overall. However, as shown in **Figure 9**, most of Karst's usage was devoted to jobs requiring one node.

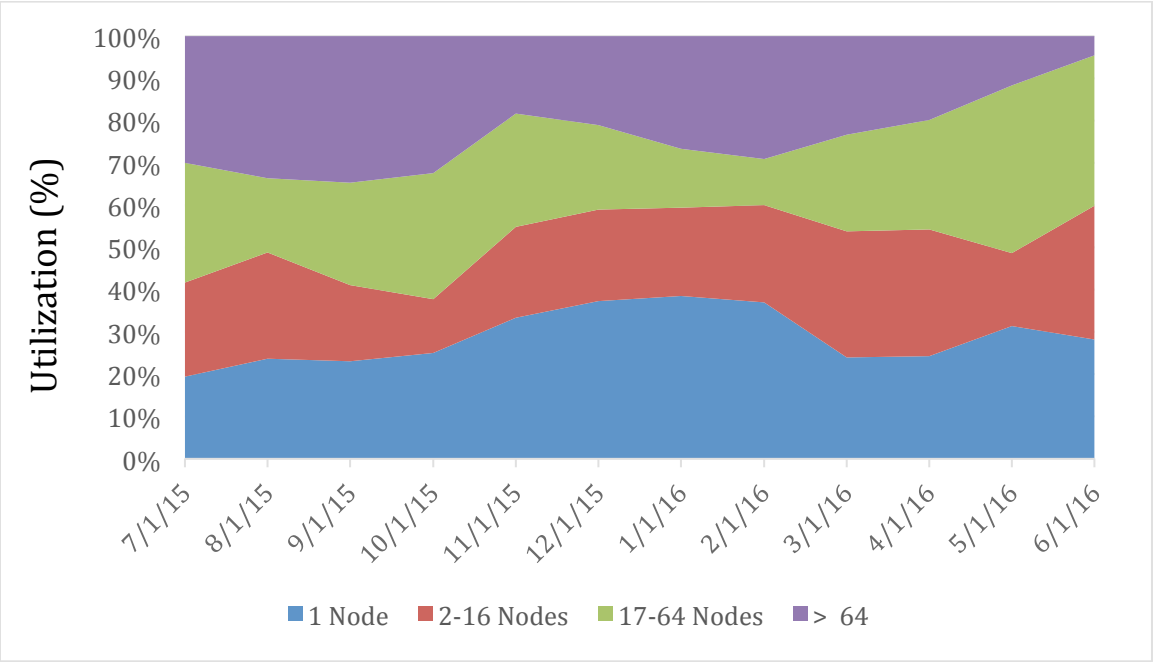


Figure 8. Percentage of Big Red II CPU utilization by job size (each node is two processors – either two CPUs or one CPU and one GPU)

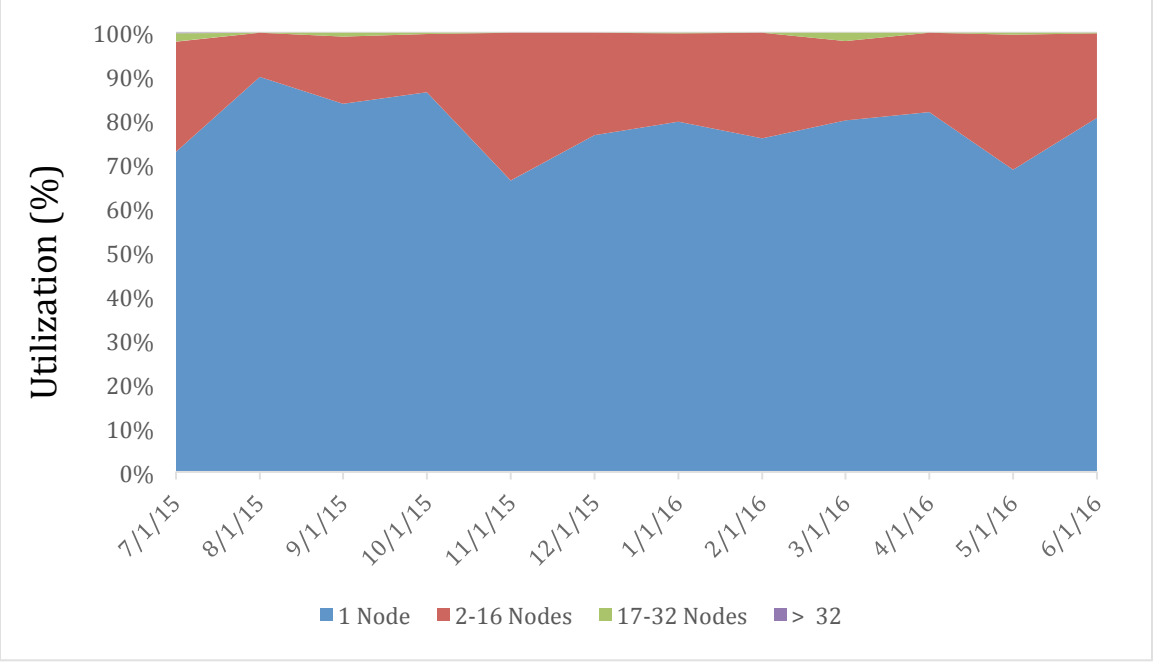


Figure 9 Karst usage for FY 2016 by degree of parallelism. Total Karst nodes: 263, for a total of 87.5 TFLOPS.

Overall, Research Technologies delivered a tremendous number of computational resources.

Table 7-B. Research Technologies delivery of resources to the IU research community – counting local cyberinfrastructure and use of national cyberinfrastructure supported by RT

System	Jobs		CPU/Resource hours	
	FY 2015	FY 2016	FY 2015	FY 2016
Big Red II	319,106	397,765	128,239,686	121,144,818
Karst	424,329	844,529	17,025,544	21,633,993
Mason	70,648	77,495	5,961,997	3,168,193
Quarry	3,814,570	Decommissioned	5,200,548	Decommissioned
XSEDE use by IU researchers	1,277	7,530	195,664	450,750
Open Science Grid use by IU researchers	25,808,916	27,164,520	26,619,180	27,010, 916
Total	30,438,846	28,491,839	183,242,619	146,397,756

7.1.1.3. Evolve IU's approaches to data and systems in ways that enable best practices across the university

- Indiana University has long been a leader in the creation and deployment of systems for data management. In 1999 then-Vice President McRobbie made the protection of IU's data assets a university priority when IU first established the highly secure tape storage system now known as the Scholarly Data Archive (SDA). The SDA enables research workflows, stores state GIS data, and serves as the permanent archive for the Media Digitization and Preservation Initiative. Table 14 below presents the RT data storage systems, purpose, and PB of storage. Figure 10 shows steady growth of the SDA usage through FY 2016.
- Key highlights in the establishment of IU's leadership in data-centric computing were two NSF Major Research Infrastructure awards: the first in in 2001 for the AVIDD project (Analysis and Visualization of Instrument-Driven Data) with then-VP McRobbie as the Principal Investigator, and the second in 2005 for the Data Capacitor advanced data management system with Craig Stewart as the Principal Investigator.
    - Deployed in 2013, IU's Data Capacitor II system has extended the capability of the original with a 5x increase in bandwidth while increasing the capacity tenfold. DC II provides high-speed short-term scratch space and mid-term project space for IU's HPC users. Figure 11 shows storage levels for DC2 over FY 2016 – storage levels have been consistently high for the entire year.
- Increasingly, "condominium computing" is being used as a means by which to better utilize computing resources. At IU, researchers (individuals, labs, departments, or schools) may purchase computational nodes and house them within the IU Bloomington Data center rather than in a separate facility (office, lab, etc.). The nodes are available to the researchers within seconds; however, the burden of securing and managing the cluster is taken on by Research Technologies. Nodes can be used by others in the IU community when not in use by the researchers, they become available to others in the IU community. Because of the benefits to the IU community, UITS hosts these nodes without charging colocation, network, or power usage fees except in

cases where researchers need a non-standard configuration or if the nodes are dedicated exclusively to the researcher. There are five research group entered into a Karst condo node agreement during FY 2016.

Table 7-C. Research Technologies data storage systems

Name	File system	Purpose	Disk (PB) total (unformatted)	Disk (PB) usable (formatted)	Tape (PB)
Geode	GPFS	Main storage system for home directories and critical files stored on disk by IU researchers	1.2	0.6 (replicated)	NA
Data Capacitor II / DC-WAN	Lustre	High speed, typically short term data storage for research data	6.47	4.85	NA
Scholarly Data Archive	HPSS	Reliable tape storage, with copies of data kept by default at IUPUI and IUB to ensure security of data	1.8	1.5	18
Totals			**9.47**	**6.95**	**18**

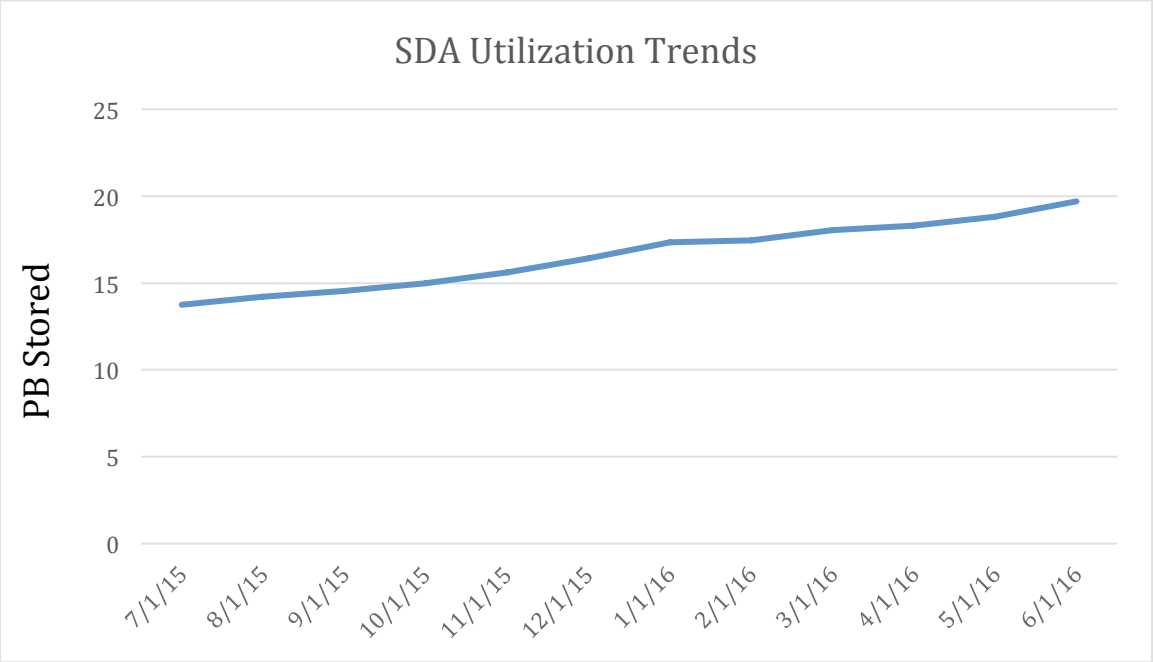


Figure 10. Utilization trends for the Scholarly Data Archive during FY 2016.

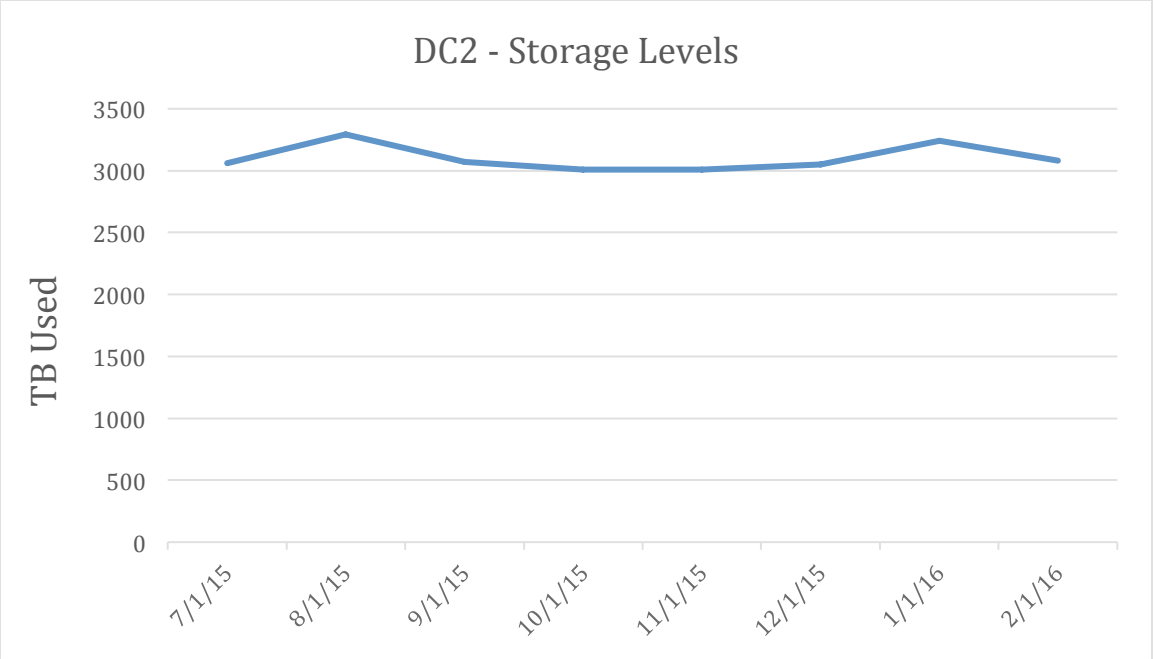


Figure 11. Data Capacitor II usage during FY 2016.

7.1.1.4. Align technology-based library and information services with physical library spaces and services

PTI supports a number of online information systems used at IU, throughout the academic community as a whole, and by the citizenry of the state of Indiana.

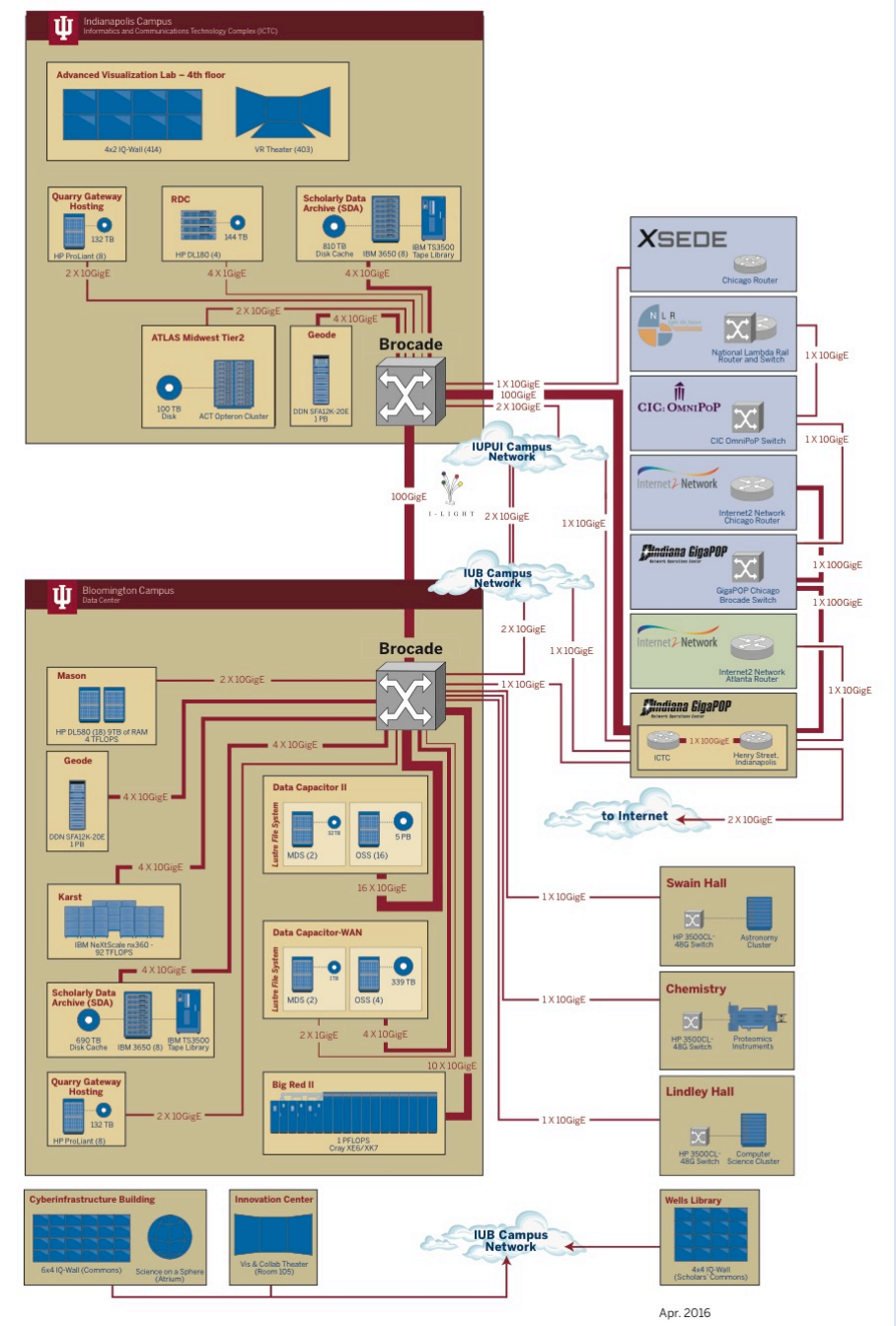
- IUScholarWorks. IUScholarWorks (scholarworks.iu.edu), operated by the IU Libraries, serves as IU's primary persistent digital repository. It is based on a front end that runs under the open source DSpace software. Its back end is the Scholarly Data Archive. IUScholarWorks is the tool IU is using to ensure that the wealth of data and information collected and generated by the IU community remains accessible to and useful for generations to come.
- **Indiana CTSI HUB.** This is the online portal for the Indiana Clinical and Translational Science Institute, and one of the most widely used data access resources within the IU clinical and translational research community.
- SEAGrid. The Science and Engineering Applications Grid (SEAGrid.org), formerly known as GridChem, is a science gateway that provides access to computational chemistry, material science, and engineering applications on IU and XSEDE computing infrastructure. SEAGrid is operated by the RT Science Gateway Group and is a tenant in the Apache Airavata-based SciGaP hosted services. During the reporting period, SEAGrid led to at least 18 scientific publications by researchers from around the world. During FY 2016, SEAGrid was the second most heavily used XSEDE gateway by numbers of jobs run and computing hours used.
- **Indiana Spatial Data Portal for GIS data**. The ISDP provides access to more than 30 terabytes of Indiana geospatial data. The ISDP provides a variety of geospatial data sets for the state, including the most recent orthophotography and lidar data commissioned by the state. Every year,

the ISDP website has thousands of visitors and enables tens of thousands of downloads (through the ISDP multi-file download interface) with total volume of data downloaded exceeding 1 TB.

- ODI-PPA. The ODI Pipeline, Portal, and Archive (ODI-PPA) is a comprehensive web-based solution that provides astronomers and WIYN Consortium members (University of Wisconsin, Indiana University, National Optical Astronomy Observatory, and the University of Missouri) with access to the One Degree Imager (ODI). The modern user interface acts as a single data access point coupled with rich computational and visualization capabilities. It supports scientists in handling complex data sets, while enhancing WIYN's scientific productivity. Most of ODI-PPA is powered by software written by (or integrated by) RT staff, and running on RT hardware including SDA. ODI-PPA has also enabled offshoot Scalable Compute Archive (SCA) projects powered by the Trident microservice and software suite, including EMC-SCA and GCS-SCA.
- **Cyberinfrastructure Gateway**. The Indiana University Cyberinfrastructure Gateway (CI Gateway) is an online portal designed to centralize information about and access to IU's advanced scholarly and artistic CI. The gateway allows users to find information on current queues, get help, see outages, find information on available software, and transfer and manage data.
- **QuakeSim/GeoGateway**. The QuakeSim science gateway (renamed GeoGateway) is the product of a long-standing collaboration between researchers at NASA Jet Propulsion Laboratory and PTI/RT. The portal gives users simple access to sophisticated InSAR datasets, GPS time series data, and forward/inversion modeling tools for comparing earthquake fault models. GeoGateway delivers over 3 TB of UAVSAR data (synthetic aperture radar data collected by aircraft) to NASA researchers.
- **UltraScan.** The UltraScan science gateway (PI Dr. Borries Demeler, University of Texas Health Science Center San Antonio) allows biophysicists to perform data analysis on analytical ultracentrifugation experiments, uncovering properties of molecules in solution. Through a collaboration with PTI/RT, this data analysis is performed on campus resources, on national cyberinfrastructure (XSEDE), and at international supercomputing centers in Germany. UltraScan is supported by the Apache Airavata-based SciGaP hosted services operated by the RT Science Gateway Group.
- **SPLInter.** The Structural Protein Ligand Interactome is a computational drug design and discovery resource for ranking molecules docked to the human proteome. The portal contains the DOPIN (Docked Proteome Interaction Network) database, which contains millions of pre-docked and pre-scored complexes from thousands of targets from the human proteome and thousands of drug-like small molecules from the NCI diversity set and other sources. SPLInter uses the Open Science Grid for docking simulations and presents visualization, scoring, and ordering information via a web portal.
- **Galaxy portal at IU**. The National Center for Genome Analysis Support provides three web-based portals that feature easy-to-use interfaces for genomics researchers to create and execute their own workflows on NCGAS systems. Using the Galaxy web portal environment, NCGAS has created Galaxy portals for IU investigators, NSF-funded life science researchers across the nation, and the Penguin On Demand system for federally-funded investigators. These provide access to the full suite of genome assembly, annotation, alignment, and other applications – as well as file transfer and transformation utilities for building genome science workflows.

Highlight: Overview of RT cyberinfrastructure

Schematic diagram of IU cyberinfrastructure showing network connections between IU and other national networks and network connections and cyberinfrastructure within IU.



7.1.1.5. Support leadership role in sustainable and energy-efficient computing

Research Technologies has been involved with the Standard Performance Evaluation Corporation for over a decade, and has contributed development effort towards High Performance Commuting benchmarks. When SPEC started to support measuring energy consumption as part of their single node benchmarks, RT submitted benchmark results for the SPEC OpenMP benchmark on the Quarry cluster. In addition, RT contributed energy measurements for running the benchmark on a regular server and within a virtualized environment. This allows for comparing not just the performance hit that users will take by running virtualized applications, but also how much additional energy a virtualized environment consumes compared to a native execution.

7.1.2. Bicentennial Priority Six: Health Sciences Research and Education to Improve the State and Nation's Health

Bicentennial plan action items to which PTI has and will contribute include:

- IUSM, and where relevant the other IU clinical schools, will continue to build research capacity in selected areas – with special focus on research in population health, cancer, cardiovascular disease, neurodegenerative diseases, and pediatrics through the Riley Children's Hospital and Foundation – in order to achieve preeminence in these areas, and to generate increased external research funding from the NIH, DOD, Patient Centered Outcomes Research Institute (PCORI), and other external sources.
- IU will invest in the infrastructure to foster collaboration in research and educational programs among the clinical schools and other academic units, leveraging the Indiana Clinical and Translational Sciences Institute (CTSI) wherever possible, in order to increase opportunities for students and researchers.

In order to support these priorities, the PTI provides several services. Here, we present operational metrics (number of databases supported, number of records in databases, and number of collaborations and studies enabled) for the operational services provided by Research Technologies to IU School of Medicine (IUSM) researchers. These services are delivered primarily via the Advanced Biomedical Information Technology Core (ABITC), a management unit of Research Technologies. ABITC was the first service outside of the school certified as an official IUSM core.

Table 7-D. Usage metrics for data resources managed and supported by ABITC and Research Technologies

Service	Number of records				Services delivered
	FY 2013	FY 2014	FY 2015	FY2016	Description
Data repository for Collaborative Initiative on Fetal Alcohol Spectrum Disorder	3,420	3,732	4,017	22,195	Number of distinct subjects (people)
	34,751	39,275	43199	69,769	Total number of database entries
	412	524	591	635	Brain images stored in CIFASD Imaging Core data repository
Data repository for National Gene Vector Biorepository and Coordinating Center	107,576	111,550	124, 588	146,484	Number of database records
Indiana CTSI HUB	4,729	6,070	7,381	7,282	Number of login (accounts) on the Indiana CTSI Hub
INResearch	1,992	3,367	4,169	5,038	Number of completed health profiles in database
REDCap	681	935	1,141	1,562	New projects using REDCap initiated by researchers
	1,362	2,297	3,438	5,000	Total projects using REDCap since 2010
CTSI Grants Management System	28	25	31	34	Grant competitions managed
	106	105	54	66	Proposals/submissions awarded
	96	121	152	176	Grant competitions managed since inception in 2009
Total Records	**153,161**	**164,634**	**184,592**	**258,241**	

Highlight: AVL and IU School of Dentistry create digital prosthodontics workflow

The Advanced Visualization Lab (AVL) and collaborated with Dr. Travis Bellicchi, a second-year maxillofacial prosthodontics resident in the IU School of Dentistry at IUPUI, to create a new digital workflow for facial prosthetics. Traditional facial prosthetics creation is a time consuming, expensive and often uncomfortable process for the patient. These older methods require casting of the patients face and hand sculpting prosthetics molds and mounting hardware. Dr. Bellicchi had the foresight to know that the process could be vastly improved by modern 3D scanning and printing. The new digital facial prosthetic workflow is quicker, more accurate, and more cost effective.

3D surface scan of a human head. The orange section renders an early prototype of the mandible.

The workflow utilizes 3D surface extraction from CT data, 3D surface scanning, and 3D printing. The process was piloted with two patient studies - an ear prosthetic and a mandible. After successful initial workflow design Dr. Bellicchi and AVL staff collaborated with Zeb Wood (IU School of Informatics and Computing), and IU student Cade Jacobs who utilized the workflow and added final artistic touches to prosthetics design for 3D printing. Initial results have gained attention of local news media. Bellicchi, Wood, and Jacobs continue to utilize the workflow and improve upon it with new patient studies.

Highlight: Using 3D imaging to identify prenatal alcohol exposure

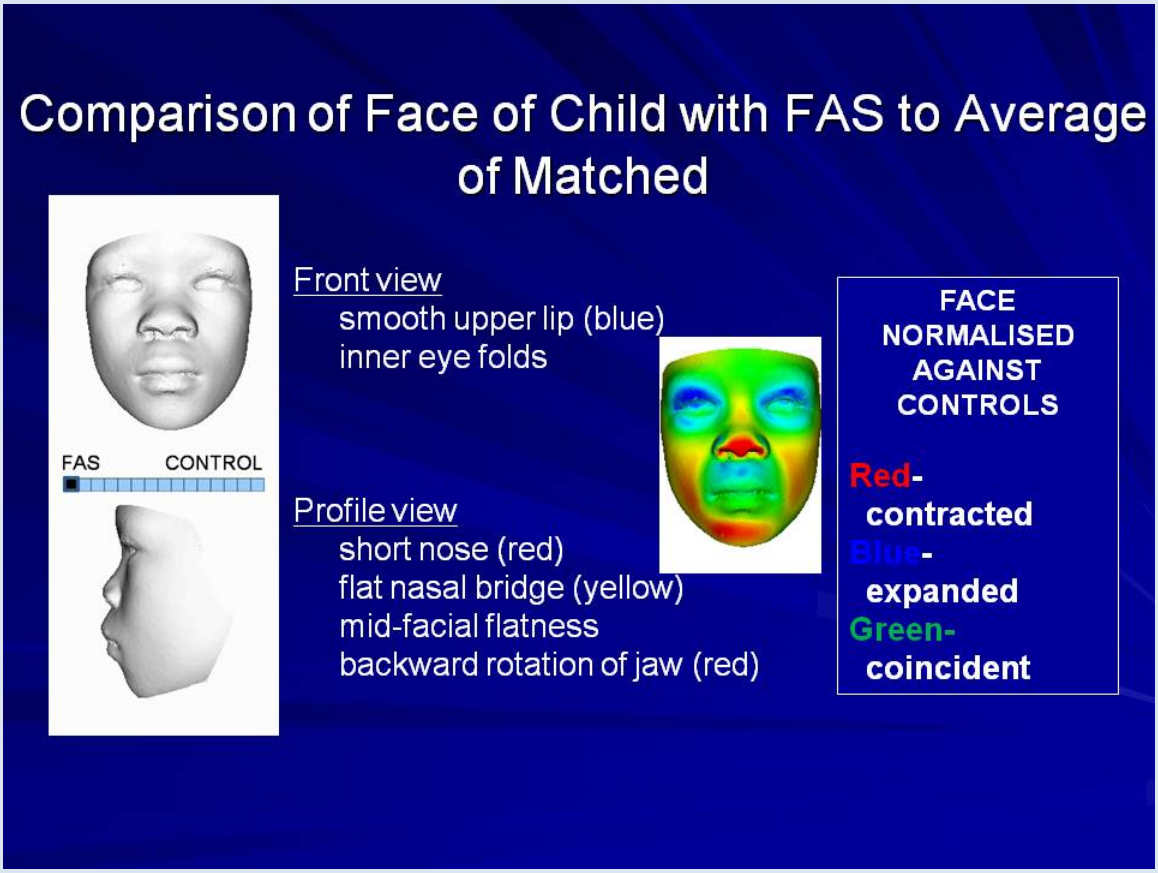


Image courtesy http://cifasd.org/research/#Foroud-Hammond

The Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD) seeks cross-cultural assessments of FASD, in order to improve clinical screening tools that provide early identification of children exposed to alcohol prenatally. The goal is to create an international collaboration of multi-disciplinary researchers using cutting-edge techniques and data to further the field of study.

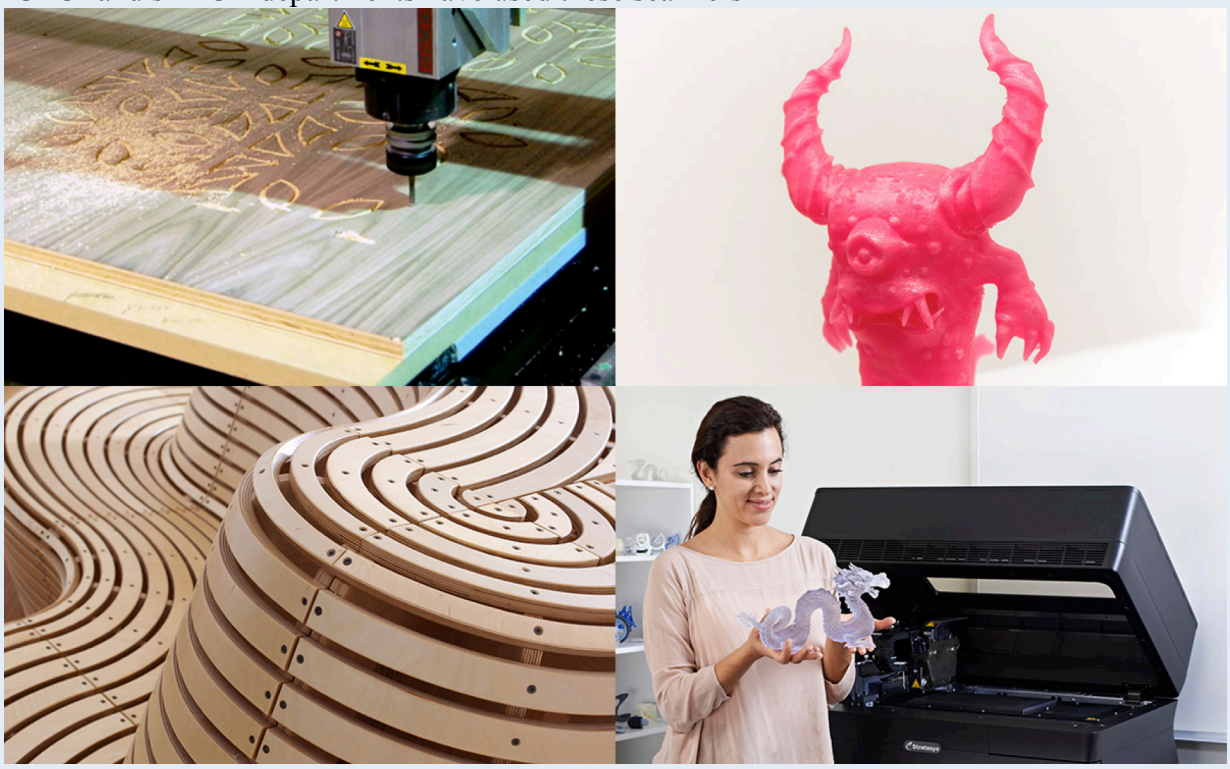
Principal Investigators Tatiana Foroud and Peter Hammond are leading a CIFASD project focused on the identification of the facial features of FASD through 3D imaging. With the aid of 3D surface scanning technologies, they hope to better understand how the face changes and how those changes correlate with other conditions related to prenatal alcohol exposure.

The UITS Advanced Visualization Lab (AVL) provides research support for this 3D facial imaging, alleviating technical challenges related to data acquisition. This support includes the deployment and technical support of six portable 3D facial scanning systems to the US and international locations like South Africa and the Ukraine.

7.1.3. BICENTENNIAL PRIORITY EIGHT: TOWARDS A CULTURE OF BUILDING AND MAKING

Highlight: RT enables Making at IU

Making and a culture of building take many forms. With support from the Advanced Visualization Lab (AVL), all IU campuses have access to two new high-resolution 3D object scanners, which can be checked out and used by members of the IUPUI community. Scanned images can then be displayed on computer monitors, or turned into physical objects via the MakerSpace or the Think It Make It Lab at the Herron School of Art. So far, at least eight IUPUI and six IUB departments have used these scanners.



Examples of projects from the Think It Make It Lab (images: Herron staff, Michelle Pemberton, Stratasys and EZ router)

Highlight: Makevention 2015

Makevention is where maker groups in Bloomington area communities come together to share the do-it-yourself spirit with each other and the community. These makers encompass a broad range of fields, including tech enthusiasts, artists, educators, crafters, hobbyists, and tinkerers. Sponsored in part by IU Research Technologies – Pervasive Technology Institute, Makevention (August 2015) attracted 1200 people with 25 groups exhibiting. Bloominglabs, a local hackerspace, helped organize the event. Jenett Tillotson and Nathan Heald from RT play a leading role in supporting Bloominglabs and organizing events such as Makevention.

One of the makers at Makevention 2015. Photo by Greg Chaney
https://www.flickr.com/gp/chained_images/FL862fhttps://www.flickr.com/gp/chained_images/FL862f

This event drew together the local community of makers of all ages. Attendees sparred with the Shire of Mynydd Seren from the Society for Creative Anachronisms, explored the myriad of projects from Bloominglabs, learned about locks from the Fraternal Order of Lock picking Sport and Columbus Key, built (and destroyed!) cup towers with Wonderlab, checked out hand-made guitars with Ivy Tech, solved mechanical puzzles with the Lilly Library, cut out jigsaw puzzles with Press Puzzles, interacted with larger than life-sized puppets from Hamster Press/Klingel-Engle, rode a virtual roller coaster with the IU Tech Showcase, stacked containers with the Quadrangles FIRST Robotics Team, drooled over the MakerMobile with Velocity Indiana and the TechShop, built rain barrels with ShareBloomington, sewed LEDs into textiles with IU Creativity Labs, folded gigantic origami figures with Discardia, browsed hand-made items from Comfort and Joy and Patricia's Wellness Arts Cafe, created videos with Level Up and the Monroe County Public Library, mixed additives into soap with Soapy Soap Company, watched the LEGO Rube Goldberg contraption and more. Everyone is having fun with these activities while also learning about science, engineering, or building through hands-on activities.

8. Appendix 1: Listing of PTI and Research Technologies Facilities and Services

1. Physical facilities

IU's cyberinfrastructure leverages the university's unusual arrangement of two major research campuses separated by 50 miles and connected by university-owned optical networks. This creates tremendous resilience in case of natural or man-made disaster, and provides an outstanding testbed for development of grid and distributed computing innovations. **Error! Reference source not found.** summarizes IU's data center facilities. IU has at present a net of 1 MW of power available to support new and expanded research cyberinfrastructure.

Table 8-A. Summary of physical facilites at IU

	Machine room total ft²	Avail. ft²	Power total	Net power avail.	Cooling capacity total (tons)	Cooling capacity avail. (tons)
ICTC	8,300	1,400	600 kW	70 kW	290	150
IUB Data Center	30,000	15,000	2.5 MW	1 MW	2750	550

1.1. IU Bloomington Data Center

The IU Bloomington Data Center (http://it.iu.edu/datacenter/) provides a highly secure and green environment for IU's largest computational and storage systems. The facility is secured with card-key access, biometric authentication, and 24x7x365 video surveillance. Only staff with systems or network administration privileges have access to the machine room. Fire suppression is provided by a double-interlock system accompanied by a Very Early Smoke Detection Apparatus (VESDA). Three circuits feed the new Data Center, travelling redundant physical paths from two different substations. Any two circuits can fully power the building.

1.2. Informatics & Communications Technology Complex

The Informatics and Communications Technology Complex (ICTC) houses IU's Data Center in Indianapolis. The ICTC is secured with card-key access and 24x7x365 video surveillance. Fire suppression is provided by a dry-pipe, pre-action sprinkler system in accordance with university risk management policy. The electrical design for the ICTC includes UPS service and generator backup for the entire facility.

1.3. Cyberinfrastructure Building

The Cyberinfrastructure Building (CIB) on the Bloomington campus opened in August 2011. Located in Technology Park East along with the IU Bloomington Data Center and the Innovation Center, the CIB houses University Information Technology Services (UITS) staff. Previously, staff were located in multiple buildings across campus. The CIB brings UITS staff together to work more efficiently and effectively than ever before.

1.4. Sustainability of physical facilities

IU Bloomington's Data Center is significantly more efficient than former facilities. The walls are made of 9,000 cubic yards of poured concrete with several sustainability features: longevity; thermal mass that decreases heating and cooling needs; recycled content; minimal waste; and regional production. The

single-story facility is surrounded by an earthen berm, offering added insulation and protection from weather events, including tornadoes up to and including category 5. The walls and berm together reduce the impact of temperature extremes and help lower heating and cooling costs and energy requirements. The mechanical equipment includes air- and water-side economizers to improve the cooling efficiency. Construction took into account federal guidance regarding sustainability and green business practices. IU achieved LEED Gold certification for the CIB and LEED Silver certification for the Innovation Center.

2. Overall structure and support of IU's advanced research cyberinfrastructure

PTI staff support all local, national, and international users of IU's research cyberinfrastructure as part of their ongoing operational responsibilities. This includes support for high performance computing systems, data storage systems, and visualization systems. Periodically, external grants and contracts help fund this support when the contract or grant determines terms of or access to such services. Otherwise, this support is made possible by IU general funds.

Online support is provided 24x7x365 by IU's award-winning Knowledge Base (kb.iu.edu). Support for security and emergency situations is provided by telephone 24x7x365 via staff at the IU GlobalNOC. In-depth support is available via email, telephone, and in-person meetings.

3. High performance computing (HPC) systems

IU has the following production high-performance computing systems.

- *Big Red II.* Big Red II is a 12-cabinet, 1-petaFLOP supercomputer from Cray Inc., installed in April 2013. Big Red II has an aggregate peak theoretical capability of just over a petaFLOP, and an aggregate RAM of 43.6 TB. Big Red II consists of 1,020 nodes total: 344 XE6 nodes, each with two 2.5 GHz AMD Abu Dhabi processors and 64 GB of memory; and 676 XK6 nodes, each with one 2.3 GHz AMD Interlagos processor and one NVIDIA K20 accelerator with 32 GB of system memory and 5 GB of GPU memory. The nodes are interconnected in a 3D Torus using Cray's Gemini interconnect, which provides 20 GB/s of bandwidth per node. This system has more bandwidth to high performance file systems like the Data Capacitor than ever before, as they connect via a low-latency InfiniBand network that provides an aggregate throughput to storage of 48 GB/s. Big Red II includes 180 TB of local spinning disk storage.
- *Karst.* Karst is a new, high-throughput cluster that includes condominium services for research groups. Karst has an aggregate peak theoretical capability of 91.5 teraFLOPS, and an aggregate RAM of 11 TB. Karst consists of 275 nodes total. Of these, three larger-memory nodes are dedicated to supporting NIH users. Karst has a 10-gigabit Ethernet interconnect. Karst includes 450 TB of local spinning disk storage. Karst features standard Linux nodes – grouping in a condominium style means nodes owned by a particular user are available for IU community use when not needed by the owner.
- *Mason.* Mason is an HP distributed, shared-memory cluster with 576 processor cores, 9 TB total memory capacity, and a peak theoretical capability of 4 TFLOPS. The compute nodes consist of 18 DL580 G7 servers, each with four 8-core Intel Xeon L7555 processors, 512 GB of memory, and a PCIe 10Gb-Ethernet adapter for high-bandwidth data transfer. The cluster includes 18 TB of local spinning disk.
- *US ATLAS Midwest Tier 2 Center* (http://mwt2.usatlasfacility.org/). The IU portion of the MWT2 facility is a heterogeneous cluster of 20 Dell 1950 servers, 56 Dell R410 servers, and 80 white-box servers, connected by a 1.0 Gbps network. This heterogeneous cluster has a total of 1,312 processor cores, 4.0 TB total memory capacity, and a peak theoretical capability of 13.6 TFLOPS. The Dell and HP compute nodes include a mix of 4-core Quad Core Xeon E5440 Processors and 6-core Intel Xeon CPU X5660 processors, with between 2 and 4 GB of memory per core. The white-box servers include a mix of Dual- and Quad-Core AMD Opteron processors. The IU MWT2 center includes 156 TB of local spinning disk storage.

- *Research Database Complex.* The Research Database Complex (RDC) is dedicated to research-related Oracle and MySQL databases and data-intensive applications based on relational databases and web applications that rely on database back ends. RDC has an aggregate peak theoretical capability of 0.3 teraFLOPS, and an aggregate RAM of 0.3 TB. RDC consists of 5 nodes total. The database-serving component of RDC consists of 4 HP DL160 servers, each with dual Intel E5620 processors, two 72 GB SAS disks, and 72 GB of memory. The RDC web serving environment is a Dell 2950 with a Quad-core Intel Xeon processor and 8 GB of memory. The RDC has a 10-gigabit Ethernet interconnect. The RDC has 144 TB of SAN-attached storage for database hosting.

Table 8-B. Summary of computation resources at IU

Name	Architecture	TFLOPS	Total RAM (TB)	Local disk (TB)
Big Red II	Cray XE6/XK6 (AMD x86-64 and NVIDIA K20)	1000.4	43.6	180
Karst	IBM NeXtScale nx360 cluster	91.5	11	450
Mason	HP DL360/580 cluster	4.0	9	18
RDC	HP DL160 database servers, Dell 2950 Web server	0.3	0.3	144
Totals		**1097.2**	**64.9**	**1056**

4. Data storage systems

In addition to the locally-attached storage listed above, IU has three major disk-based file systems and one archival storage system that serve local and remote users. These include:

- *Geode.* Geode is a new disk-based filesystem with a capacity of 1.2 PB (600 TB usable due to replication) that allows for group collaboration via file sharing. Users have a highly flexible system for granting access to files, and the underlying GPFS technology used for the system replicates all data to both IU Bloomington and IUPUI for a highly available system architecture. Researchers can request dedicated project space for each project requiring dedicated storage and collaboration. Users can access files from their desktops (CIFS) and via SFTP. This system is part of the replacement for the Research File System (RFS) based upon OpenAFS and the previous HPC home directory solution, which had been served by a NetApp NAS. Furthermore, Geode is providing Condo storage, which allows research departments to purchase large allocations of storage (beginning at 100 TB) to store data for projects across their departments.
- *The IU Data Capacitor II.* The Data Capacitor II is a high-speed/high-bandwidth Lustre storage system that serves the high performance computing systems at IU Bloomington. Installed in February 2013, DCII includes a 5 PB Lustre file system (4 PB usable) from Data Direct Networks (DDN), with an aggregate 48 GB/s of data I/O capability. DCII includes 16 Lustre object storage servers, two Lustre metadata servers, and eight Lustre routers, all connected via full-data-rate (FDR) InfiniBand to two DDN SFA12000 storage controllers. The Lustre routers also have 10-gigabit Ethernet connections to allow systems without InfiniBand to access the system. The two metadata servers are connected to a DDN SFA6620 storage controller.
- *The Data Capacitor Wide Area Network (DC-WAN)* file system is a high-speed/high-bandwidth Lustre storage system for research computing that serves all IU campuses and other sites throughout the country, primarily by wide area network (remote) Lustre file system mounts. DC-WAN has a total formatted capacity of 1.1 PB, with 40 Gbps maximum I/O. DC-WAN consists of Dell 2950 servers running the Lustre file system. DC-WAN has four servers equipped with 10-gigabit Ethernet cards for object storage, and two that use Gigabit Ethernet for Lustre metadata. DC-WAN can map remote

users to local users, allowing machines with heterogeneous namespaces to communicate seamlessly. DC-WAN currently serves a legacy role within XSEDE for 40 XSEDE project allocations that require DC-WAN's capabilities for file storage and long distance accessibility. These are projects that have not found other XSEDE services that meet their needs. IU provides wide area file system connections for over 10 collaborators and facilities.

- *IU's Scholarly Data Archive (SDA)*. SDA uses High Performance Storage System (HPSS) software to make available to IU researchers a total storage capacity exceeding 18 PB. Data is written to a fast, front-end disk cache and migrated over time to IBM TS3500 tape libraries on the Indianapolis and Bloomington campuses. Data written to IU's HPSS system are copied simultaneously to both locations, providing highly reliable disaster protection. Users can access data over the network from central research systems or from personal workstations, using SFTP, HSI/Htar, CIFS, and HTTP. The default allowance is 50 TB of mirrored data, with additional space provided upon request. SDA stores and provides access to data for the IUScholarWorks Repository (http://scholarworks.iu.edu), a document and data archiving system created using DSpace software.

Table 8-C. Summary of data storage resources available at IU

Name	File system	Disk PB unformatted	Disk (PB) usable (formatted)	Tape (PB)
Geode	GPFS	1.2	0.6	NA
Data Capacitor II / DC-WAN	Lustre	6.47	4.85	NA
Scholarly Data Archive	HPSS	1.8	1.5	18
Totals		**9.47**	**6.95**	**18**

4.1. Backup and replication within IU Storage Systems

The backup and/or data replication procedures for IU storage systems are as follows:

- *The Research File System.* RFS is backed up nightly to the SDA and saves versions for at least the previous seven days, seven weeks, and two months. While users must request a restore of one of these versions, the previous day's version of each of the user's files is immediately accessible in the one-day backup directory in that user's account.
- *The IU Data Capacitor II and DC-WAN.* Data stored on the Data Capacitor II and DC-WAN are not backed up automatically. This system was designed primarily for short-term data storage. However, data from the Data Capacitor can easily be transferred to the SDA from any of IU's compute resources, so replica copies may easily be maintained.
- *IU's Scholarly Data Archive (SDA)*. By default, data stored within the IU Scholarly Data Archive are stored in duplicate copies – one in the tape silo at IU Bloomington, and one in the tape silo at IUPUI in Indianapolis. User data is not backed up to other external systems. The HPSS metadata specifying which tapes contain any given file is backed up continuously; multiple copies exist in Indianapolis and Bloomington.

As noted in Section 7, the system security and its documentation are in compliance with NIST 800 Security Standards.

4.2. Facilities for handling sensitive data

IU has put in place appropriate administrative, technical, and physical controls to protect data in accordance with the HIPAA security rule. Electronic Personal Health Information may be stored on all of the HPC and storage facilities described in this document.

4.3. Services lists and disaster recovery planning

IU has a written disaster recovery plan for every service and system it provides, which is by definition an experimental facility. (See a full list of services at: https://webdb.iu.edu/uitsfs/scripts/abc/reports/web_files/0910/RCQS/Basic/RCQS_09-10_UA_BASIC.pdf) IU also has a contract in place for use of an off-site disaster recovery facility in case of a disaster affecting one or more of IU's campuses. If a disaster strikes one core campus (IUPUI or IU Bloomington), the disaster recovery plans call for restoring service at whichever core campus remains operational. Plans are also in place for service recovery if a disaster strikes both core campuses simultaneously.

5. Networking

The basic logical structure of IU data networks can be subdivided into three primary components: connections between high-speed research networks and commodity Internet to IU, the IU research network, and the IU enterprise network. These network connections and major features of the IU cyberinfrastructure are shown in Section 7 (see page 41). The IU research network meets ESNet's definition of a Science DMZ and, by placing research systems outside the general campus network, has operated as such since 2004.

The primary connection between IU and national research networks is a 100 Gbps network link from Internet2 to the Indiana GigaPOP in Indianapolis. The Indiana GigaPOP is a collaborative facility operated by the IU GlobalNOC on behalf of collaborating partners Ball State University, Indiana University, Purdue University, and the University of Notre Dame. IU was the first site to connect to Internet2 at 100 Gbps as part of the Monon100 project. In the first half 2013, IU's 100 Gbps was extended from Indianapolis to the main campus in Bloomington. IU also has a dedicated 10 Gbps connection to the XSEDE network. In 2013, the Indiana GigaPOP connection to the CIC OmniPOP was upgraded to 100 Gbps. For redundancy, the GigaPOP also maintains four 10 Gbps redundant and physically isolated connections to commodity Internet.

The IU Research Network uses the 100 Gbps connection installed as part of the Monon100 project as its backbone from the Indiana GigaPOP to the IUPUI campus in the Informatics and Communications Technology Complex building, and from there to Bloomington and the IU Bloomington Data Center (both physical facilities described in section 1). Seventy percent of the capacity of these links is dedicated to research use. Within both data centers on the IUPUI and IU Bloomington campuses, Brocade switches function as "machine room backplanes." These switches provide connections among cyberinfrastructure systems within the data centers. Through them, all cyberinfrastructure systems in Indianapolis and Bloomington are interconnected.

IU maintains a separate enterprise network for business and general academic use. This enterprise network peers with the research network in Indianapolis and Bloomington to provide users with redundant connectivity. The IU research network also has a redundant path to the commodity Internet via the enterprise network, in case connectivity between Indianapolis and Bloomington is interrupted.

6. Advanced visualization facilities

The IU Advanced Visualization Laboratory (AVL) serves as a university-wide resource for visualization, virtual reality, advanced graphics, and visual telecollaboration for researchers, educators, students, and artists in all departments on all campuses. AVL has eight full-time staff and can host graduate students for extended projects. The AVL maintains and operates a number of advanced visualization resources,

including but not limited to ultra-high-resolution displays, virtual reality environments, 3D cameras, scanners, spatial input systems, and haptic feedback devices.

Table 8-D. Visualizations facilities for the AVL

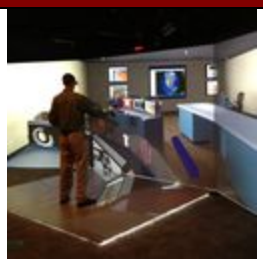
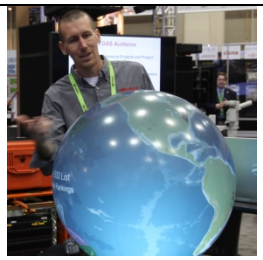
Image	Description
	*The Virtual Reality Theater* is a bright, high-resolution, immersive virtual reality technology resource suitable for individual and group use. The Theater is reconfigurable and driven by workstation computers running either Windows or Linux.
	*The Visualization & Collaboration Theater*, a three-screen, front-projected display, offers either stereoscopic or monoscopic high-definition visualization and presentation capabilities for up to 60 people.
	*Science on a Sphere (SOS)* is a spherical display developed by NOAA that blends four high-resolution projectors to create a seamless image on a globe nearly 6' in diameter. IU's SOS is located in the atrium of the Cyberinfrastructure Building (CIB) on the Bloomington campus, and can be used to display a variety of scientific and information visualizations as well as digital art and other interactive experiences.
	*Puffersphere* is IU's newest spherical display. It features a smaller footprint than SOS – which allows it to more easily move between buildings or campuses – as well as a multi-touch interface to facilitate new forms of collaboration and interactive exploration.
	*The Ultra-High Resolution Display Wall* consists of eight high-resolution projection cubes totaling 15.3 million pixels. The Display Wall is capable of receiving input from multiple sources simultaneously, making it ideal for teleconferencing, group collaborations, and/or multiple highly advanced visualization applications. Like the IQ-Wall described below, it is driven by a single computer.

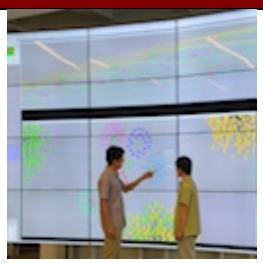
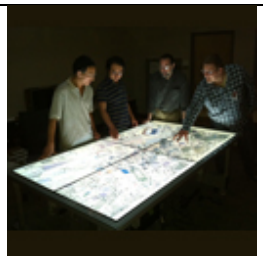
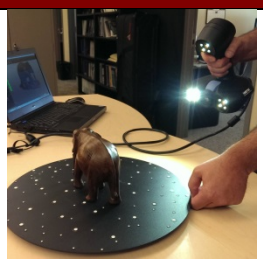
Image	Description
	*The IQ-Wall* is an AVL design that ties thin, energy-efficient, flat-screen monitors into configurations that meet the requirements of the users and the space. These walls are driven by a single Windows computer. The premier unit is located in the lobby of the Cyberinfrastructure Building (CIB) at IU Bloomington.
	*The IQ-Table* is a 55" monitor equipped with 32-point multi-touch capabilities, which make it ideal for lobbies, libraries, or off-site exhibits. Its built-in computer and shipping case make it easy to ship, set up, and tear down.
	*The IQ-Tilt* features four monitors tiled together in a 2x2 configuration. This nearly 100" display is treated as one Windows desktop driven by a single computer, and is multi-touch enabled. Its name comes from the fact that this display pivots on an axis, and can be reconfigured in fewer than 10 minutes into either a horizontal table or a vertical wall.
	*The IQ-Force* features a tabletop device that combines stereographic rendering with physical force and tactile feedback. When users can see their own hands in the same physical vicinity as the virtual simulation, they experience a more natural sense of scale, orientation, and augmentation.
	*The IQ-Station* is a low-cost stereoscopic display with optional user interface components including touchscreens and/or input device tracking. A fully equipped Station is most suitable for scientific research groups; simplified versions are excellent for stereoscopic screening and production, and for conferences and outreach events.
	*Head Mounted Displays* are available through AVL, along with comprehensive software workflows for creating and deploying virtual reality experiences. AVL's primary HMD is the Oculus Rift, but other mobile VR platforms are available. AVL offers training and assistance as well as limited licensing opportunities for the Unity Engine. This combination of Rift hardware and Unity software lowers the barrier of entry to creating immersive experiences.

Image	Description
	*3D Scanning and Digital Object Preparation* are areas in which AVL staff have knowledge and workflows broadly related to object and environment scanning. AVL owns and operates multiple 3D scanners, which can be used in AVL facilities or loaned out. To complement the scanning process and preparation for digital fabrication (i.e., 3D printing), AVL staff are also well versed in converting and manipulating all types of digital 3D models.
	*Advanced Media Capture Equipment* maintained by AVL includes a high-end, professional stereoscopic camera rig capable of filming authentic productions. Video captured with this rig can be edited using software available to all IU students, faculty, and staff. AVL's Gigapan robot can be used to automatically capture multiple 2D photographs of a scene or environment. Once the individual photos have been composited together, users can interact with the final panoramic image via web browsers. AVL also has THETA cameras capable of capturing 360-degree images and videos with a single click. These images and videos look great in an immersive, head-mounted display.

7. Federal systems security policy and federal funding agency policy compliance

The IU high-performance computing and storage systems described here are managed and administered in ways that meet National Institute of Standards and Technology (NIST) 800 security standards. OVPIT and UITS comply with the NIH Grants Policy Statement.

8. Central information technology organizations

The Indiana University Office of the Vice President for Information Technology (OVPIT) and University Information Technology Services (UITS) are responsible for delivery of core information technology and cyberinfrastructure services and support. OVPIT and UITS collectively have an annual budget of more than $110,000,000 and employ more than 700 full-time staff members. The Pervasive Technology Institute (PTI) – a collaborative effort of OVPIT, UITS, the IU School of Informatics, and the Maurer School of Law – is IU's flagship resource for information technology research, development, and delivery.

9. Appendix 2: EOT activities

Date	Education, Outreach, and Training Event Title	Conference Name/Location	Description	Total Attendees
7/7/2015	Kia Ekbia visit to AVL	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour and discussion and demo lead by AVL staff.	1
7/8/2015	CASC/HPC-SIG Joint Oxford Workshop Report: Towards Sustainable National Research Computing Centers and a Computational Workforce	CASC/UK HPC-SIG Workshop / Oxford University e-Research Centre	Networking workshop between members of CASC and the UK HP-SIG that led to report	18
7/13/2015	AVL Tour for SOIC Web Dev Summer Camp	IUPUI - Informatics & Communications Technology Complex AVL	An informal tour and discussion and demo lead by AVL staff.	7
7/13/2015	Ready, Set, Robots! 1/2 day workshop for Jr. High students	Minority Engineering Advancement Program (MEAP) - IUPUI	Students programmed a mission to Mars using LEGO and the Next program.	25
7/15/2015	AVL Tour for SOIC Pokemon Summer Camp	IUPUI - Informatics & Communications Technology Complex AVL	An informal tour and discussion and demo lead by AVL staff.	19
7/16/2015	AVL Tour for Leon Nowlin VFX Alumni from Informatics	IUPUI - Informatics & Communications Technology Complex AVL	An informal tour and discussion and demo lead by AVL staff.	1
7/16/2015	Ready, Set, Robots! Advanced Summer Camp	IUB - Cyberinfrastructure Building	Jr. high and high school students learn computer programming including binary code.	15
7/22/2015	AVL Tour for First Baptist Church Youth Camp Tour	IUPUI - Informatics & Communications Technology Complex AVL	An informal tour and discussion and demo lead by AVL staff.	18
7/23/2015	AVL Tour for Mobile App Development Summer Camp	IUPUI - Informatics & Communications Technology Complex AVL	An informal tour and discussion and demo lead by AVL staff.	9
7/27/2015	HILT evening reception	HILT workshop	open evening reception for workshop group	60
7/27/2015	XCBC & XNIT using the LittleFe and Limulus HPC200	XSEDE15 / St. Louis, MO	Tutorial on XCBC/XNIT using small platforms	4
7/27/2015	Educational activity 1/2 day workshop for grade school	Minority Engineering Advancement Program (MEAP) - IUPUI	Students learned about hardware and networking.	25
7/27/2015	XSEDE Value Added Cost Avoidance and Return on Investment	XSEDE15 / St. Louis, MO	Lightning talk by Craig A. Stewart	280
7/28/2015	Summer camp	Kent State CS department summer camp for Big Data and Cyberinfrastructure | IUB - Cyberinfrastructure Building Innovation Center	Host camp for undergraduate students.	31
7/28/2015	XSEDE Value Added Cost Avoidance and Return on Investment	XSEDE15 / St. Louis, MO	Presentation by Craig A. Stewart	40
7/28/2015	Jetstream - A self-provisioned scalable science and engineering cloud environment	XSEDE15 / St. Louis, MO	Presentation by Craig A. Stewart	35
7/29/2015	Jetstream Overview - New and emerging US cyberifrastructure resources	XSEDE15 / St. Louis, MO	Presentation as part of a plenary panel about Jetstream by Craig Stewart	300
8/3/2015	Educational Activity for Scholarship winners	USArray Short Course / IUB - Lindley Hall and IMU	Robert Ping organized the workshop.	36
8/5/2015	Parallel Programming	USArray Short Course / IUB - Lindley Hall and IMU	Ray presented a 1/2 workshop.	36
8/5/2015	XSEDE Value Added Cost Avoidance and Return on Investment and Building	UITS RT/PTI Seminar / IUB and IUPUI	Stewart and Knepper presented on XSEDE initiatives.	30

Date	Education, Outreach, and Training Event Title	Conference Name/Location	Description	Total Attendees
	XSEDE-like clusters on campus with XCBC and XNIT			
8/12/2015	Introductory workshop on the use of Galaxy for bioinformatics	Bioinformatics Clinic / IUB - Lindley Hall	NCGAS staff taught a workshop	10
8/19/2015	An introduction to UITS Research Technologies services	Graduate Program Student Orientation (GPSO) Fair / IUB - Frangipani Room	Presented materials at a table during the fair.	21
8/24/2015	Collaborative meeting	IU / Broad Trinity Development Workshop / IUB - Cyberinfrastructure Building	An on-going collaboration with Broad Institute researchers	8
9/3/2015	Short AVL tour for Bloomington UITS finance staff member	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour for a staff member based on the IUB campus.	1
9/3/2015	Intel Xeon Phi Workshop	IUB - Cyberinfrastructure Building	IU allowed Intel to host a workshop in CIB.	60
9/8/2015	Campus Bridging: Reducing Obstacles on the Path to Big Answers	IEEE Cluster 2015 / Chicago IL	CBRI organized a workshop presented to attendees of the IEEE Cluster conference	16
9/8/2015	S4ES: Introductory Parallel Programming for Supercomputers	IUPUI - ICTC IT 121	Taught introductory parallel programming	18
9/10/2015	S4ES: Introductory Parallel Programming for Supercomputers	IUB - Wells Library W144	Taught introductory parallel programming	19
9/16/2015	Oculus Rift Demonstration	SOIC Lawn Party / IUPUI ICTC court yard	Display of the oculus rift virtual reality system for students to try the system and talk to AVL staff.	150
9/17/2015	Educational Activity	IUB - Herman B Wells Library Scholars' Commons IQ-Wall	AVL staff introduced AST-A 780 students to AVL services and the IQ-Wall	17
9/22/2015	S4ES: Faster Work Safer Storage	IU-Bloomington and IUPUI	Introduction to HPC workshop	27
9/25/2015	AVL Tour for Tim Scales and guests	IUPUI - Informatics & Communications Technology Complex AVL	An informal tour and discussion and demo lead by AVL staff.	4
9/30/2015	AVL Tour for INFO-I 100 Course	IUPUI - Informatics & Communications Technology Complex AVL	An informal tour and discussion and demo lead by AVL staff.	16
10/2/2015	Educational Activity	IUB - Cyberinfrastructure Building and Innovation Center	AVL staff hosted a potential swimming recruit	2
10/3/2015	Ready, Set, Robots! and Scribble Bots	Celebrate Science Indiana / Indiana State Fairgrounds - Indianapolis, IN	Had several activities for K-12 and their families for STEM outreach.	350
10/6/2015	AVL for new informatics students	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour for new Informatics students based on the IUPUI campus.	35
10/7/2015	PTI, UITS and their role in the IU Grand Challenge Initiative	IUB - Cyberinfrastructure Building IQ-Wall	RT/PTI Seminar /	657
10/9/2015	AVL Tour for high school students	Upward Bound Health Career Day / IUPUI - Informatics & Communications Technology Complex AVL	Informal tour for students considering bioinformatics as their course of study.	7
10/12/2015	AVL Tour for N500 Informatics Graduate Class	IUPUI - Informatics & Communications Technology Complex AVL	An informal tour and discussion and demo lead by AVL staff.	9
10/13/2015	AVL Tour	IUPUI - Informatics & Communications Technology Complex AVL	#145459: EOT: Impromptu Student Tour	1
10/19/2015	Introduction to R	IUB - Herman B Wells Library	Guest Lecture in Corpus Linguistics Class for Markus Dickinson	30
10/20/2015	Jetstream - a national science and engineering cloud	Statewide IT Conference / IUB - Indiana Memorial Union	Presentation about Jetstream (given by Jeremy Fischer)	40
10/20/2015	Cyberinfrastructure for Research: New Trends and	University of Vermont, Burlington VT	Stewart C.A. 2015. Presentation: http://hdl.handle.net/2022/20414	40

Date	Education, Outreach, and Training Event Title	Conference Name/Location	Description	Total Attendees
	Tools.			
10/21/2015	Cyberinfrastructure for Research: New Trends and Tools.	University of Michigan	Stewart Craig A. Presentation: http://hdl.handle.net/2022/20445	30
10/22/2015	Cyberinfrastructure for Research: From campus growth to national trends.	Cyberinfrastructure Days / Michigan State University	Stewart Craig A. 2015. Presentation: http://hdl.han	200
10/27/2015	Academic Medical Center Pre-Educause Meeting Tour	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour for staff members and affiliates.	19
10/28/2015	AVL Tour for Informatics 101 class	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour for new Informatics students based on the IUPUI campus.	18
10/29/2015	IDEW (Informatics' Diversity Enhanced Workforce) High School Students Tours	IUPUI - Informatics & Communications Technology Complex AVL	Tour sessions and discussions lead by AVL and Informatics staff.	41
10/29/2015	pti/rt talk			--
11/2/2015	Jetstream: A new national research and education cloud	Southern Partnership in Advanced Networking (SPAN) - Workshop 2 / Huntsville AL	Workshop for southern universities supporting research and teaching activities	44
11/2/2015	S4ES: Gateways for Science Apps	IUB - Cyberinfrastructure Building and IUPUI - ICTC	First workshop at IU as part of Supercomputing for Everyone Series for Science Gateways.	10
11/4/2015	GIS Day at IU 2015	IUB - Herman B Wells Library	Annual event focused on geospatial technologies	500
11/4/2015	Introducing Graphical Interface to IU's Supercomputers	IUB - Cyberinfrastructure Building IQ-Wall	RT/PTI Seminar / Abhinav Thota introduced the Karst Desktop app.	32
11/5/2015	S4ES: Introductory Parallel Programming for Supercomputers	IUB - Herman B Wells Library W144	Taught introductory parallel programming	14
11/9/2015	AVL Tour	JagDays / IUPUI - Informatics & Communications Technology Complex AVL	An informal tour of the AVL for new students and parents.	38
11/9/2015	Selling condos when the rent is free	Supercomputing 2015 / Austin TX USA	Presented IU condo model in Lenovo Booth	12
11/10/2015	3D filmmaking	IUB - Radio/TV Building	Presentation on 3D filmmaking to Telecommunications majors	67
11/10/2015	Preparing sound for digital cinema	IUB - Jacobs School of Music	Presentation by AVL staff	12
11/11/2015	AVL Tour for Kokomo/Informatics K-12	IUPUI - Informatics & Communications Technology Complex AVL	EOT tour	30
11/11/2015	Various topics on visualization	VIsTech Workshop in conjunction with SC15 / Denver, CO	Accepted workshop led by AVL	154
11/17/2015	Demonstration of Jet Stream	The International Conference for High Performance Computing Networking Storage and Analysis / Denver, CO	Demonstration of Jetstream to conference attendees.	15
11/17/2015	Indiana University Innovations and Demonstrations	The International Conference for High Performance Computing Networking Storage and Analysis / Denver, CO	Large booth at major tech conference	1200
12/2/2015	Experiences in Higher Education with SOS	2015 SOS Users Collaborative Network Meeting / Portland OR	AVL staff presented to an SOS audience about our experience with SOS.	20
12/2/2015	Research data sharing without barriers - the Research Data Alliance and IU	IUB - Cyberinfrastructure Building IQ-Wall	RT/PTI Seminar /	42
12/3/2015	Raytheon Informatics meeting and tour	IUPUI - ICTC Advanced Visualization Lab	A discussion of the technology and research being used/done by Informatics in the AVL	5
12/3/2015	Cluster administration for non-	On the web	CBRI staff presented about cluster	15

Date	Education, Outreach, and Training Event Title	Conference Name/Location	Description	Total Attendees
	XSEDE Resources webinar		computing administration resources provided by XSEDE for non-XSEDE resources	
12/8/2015	AVL Tour for Tech 104	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour and discussion and demo lead by AVL staff.	12
12/14/2015	AVL Tour for Medical Library	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour and discussion and demo lead by AVL staff.	6
12/16/2015	AVL Tour for Mangilal Agarwal and guests	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour and discussion and demo lead by AVL staff.	4
1/6/2016	Overview of Jetstream for Earth Sciences	2016 ESIP Winter Meeting / Washington DC	Presentation to researchers and IT staff	20
1/7/2016	AVL Tour for new IUPUI Herron VCD faculty	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour and discussion and demo lead by AVL staff.	6
1/11/2016	ALL Tour for new SoIC staff	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour and discussion and demo lead by AVL staff.	3
1/11/2016	I400/I590 Informatics course on Virtual Reality	IUB -- Lindley Hall 008 (also use of Scholars' Commons and IUIC 105)	Undergraduate/Graduate combined course on virtual reality	30
1/20/2016	AVL Tour for N444 SOIC Stereoscopic Production Class	IUPUI - Informatics & Communications Technology Complex AVL	Demonstration and discussion about how stereoscopoic displays work	6
1/22/2016	Ten Easy Ways to Use the IQ-Wall	IUB - Herman B Wells Library Scholars' Commons IQ-Wall	CyberDH Spring Workshop Series	4
1/27/2016	Keynote talk: Exascale on what dimension and why?	SPPEXA Annual Program Meeting / Munich Germany	Keynote presentation by Craig A. Stewart	161
1/29/2016	R: Basics for Humanists	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	25
2/3/2016	RT services, systems, and tools	IUB - CIB	Presentation to UITS Community Partners	5
2/5/2016	R: Intermediate Scrips for Humanists	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	9
2/9/2016	Cluster Management for Non-XSEDE Systems	Presented online only	Webinar showcasing system administration resources for research-oriented clusters	31
2/12/2016	Introduction to Scalar	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	2
2/15/2016	Bring your group projects to the next level with Github:IU	IUB - Herman B Wells Library	CBRI presentation highlighting features of github:iu that would be useful for group class projects	1
2/16/2016	pti/rt talk Grimshaw			42
2/16/2016	Figuratively Speaking PIPES Research Demo Session	Electronic Imaging 2016 Conference / San Francisco, CA	An art and technology demonstration session .	130
2/16/2016	Digital Arts class presentation	IUB -- Fine Arts	Presented information on the Usage of Unity 3D software to course	15
2/17/2016	Faster Work Safer Storage February 17 and 18 2016	Faster Work Safer Storage Bloomington and Indianapolis	Introduction to HPC workshop	16
2/18/2016	Creating Virtual Worlds with Unity - Workshop	Electronic Imaging 2016 Engineering Reality of Virtual Reality / San Francisco	A dissemination of how the AVL supports clients using VR in their research using the Unity Game Engine as a VR authoring tool.	22
2/19/2016	3D Object Creation with Photogrammetry	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	7
2/22/2016	JagDays AVL Tour	IUPUI - Informatics & Communications Technology Complex AVL	Tour and demonstration of the technologies in the AVL.	29
2/22/2016	AVL Tour for N502 Informatics Graduate Students	IUPUI - Informatics & Communications Technology Complex AVL	Tour and technology demonstration.	9
2/25/2016	Galaxy workshop	IUB - CISAB building - 409 N. Park Ave	A hands-on walk-through of running RNA-Seq data through the IU galaxy	11

Date	Education, Outreach, and Training Event Title	Conference Name/Location	Description	Total Attendees
			interface.	
2/26/2016	360 Degree Panoramas and Open Source 3D Viewer	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	8
2/27/2016	Catalyst Presentation	Catalyst	Informatics event for Undergraduate networking	50
3/1/2016	UISO/UIPO Chalk Talk March-2016	IUPUI - Informatics & Communications Technology Complex	Discussion with UISO and UIPO on the Jetstream project	13
3/2/2016	AVL Tour and discussion with Creative Works Inc. Visitors	IUPUI - Informatics & Communications Technology Complex AVL	Informal AVL systems tour and high level discussion of tech.	2
3/3/2016	The Oculus Rift	Student recruitment visit / IUPUI AVL	Abbeiviated demonstration of the Oculus Rift	3
3/3/2016	Campus Bridging: Solutions for IT Staff who Support Research	EDUCAUSE Connect 2016 / Denver CO	Campus bridging staff presented a poster detailing system administration resources aimed at supporting research	8
3/4/2016	R for Humanists	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	10
3/5/2016	IUPUI Jagathon 2016	IUPUI Jagathon 2016 / IUPUI - Informatics & Communications Technology Complex AVL	IUPUI Jagathon 2016	100
3/8/2016	AVL tour for digital humanities group from Earlham College	IUPUI - Informatics & Communications Technology Complex AVL	AVL tour for digital humanities group from Earlham College	8
3/8/2016	PIPES demonstration for guest from IUPUI Office of Enrollment	IUPUI - Informatics & Communications Technology Complex AVL	Demonstration of the PIPES VR tour in the IUPUI AVL.	2
3/8/2016	Jetstream presentation	Science Gateways Architecture INFO-590 class / IUB - School of Informatics and Computing	HPS staff described Jetstream's hardware and software architecture use cases and operational aspects of a functioning cloud system	15
3/9/2016	The Apache Way with Rich Bowen	RT/PTI Seminar / IUB - Cyberinfrastructure Building IQ-Wall		38
3/9/2016	OpenStack and Apache with Rich Bowen	RT/PTI Seminar / IUB - School of Informatics and Computing		22
3/9/2016	Jetstream overview	SPAN Workshop 3 / University of Central Florida Orlando, FL	Overview of Jetstream for southeastern US IT and research staff	42
3/9/2016	Tour	IUB Datacenter	Led a tour for guests from RedHat & OpenStack	5
3/11/2016	Jetstream Architecture	IU's Security and Policy offices' monthly ChalkTalk / IUPUI - Informatics & Communications Technology Complex	HPS staff presented an architectural description of Jetstream.	13
3/11/2016	Partners in Education student group visit	IUB - CIB Data Center Innovation Center	Hosting middle school students with School of Education Partners in Education group.	23
3/14/2016	AVL facilities tour for Informatics phd Student Eric Vorm and Lawrence Walter	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour for a new graduate student and his collaborator in Informatics	2
3/14/2016	Harshman Middle School AVL Tour	IUPUI - Informatics & Communications Technology Complex AVL	Harshman Middle School students as guests of the School of Informatics	35
3/17/2016	AVL Tour for 2017 New Student Visit SOIC Guests	IUPUI - Informatics & Communications Technology Complex AVL	AVL tour for a new student of Informatics and their family.	4
3/18/2016	AVL Tour for Class Tech Services	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour of AVL systems and projects	8
3/23/2016	Using R for Social Media Mining	IUB - Herman B Wells Library	Guest Lecture for Social Media Mining Class for Muhammad Abdul-Mageed	10
3/24/2016	AVL Tour for SOIC guests	Decision Day event / IUPUI	Informal facilities tour for SOIC guests	32

Date	Education, Outreach, and Training Event Title	Conference Name/Location	Description	Total Attendees
			attending Decision Day event	
3/24/2016	RT services, systems, and tools	Hutton Scholars' Interview Day / IUB - Hutton Honors College	Presentations about UITS RT	15
3/25/2016	R: Basics for Humanists	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	12
3/28/2016	RT services, systems, and tools	Hutton Scholars' Interview Day / IUB - Hutton Honors College	Presentations about UITS RT	15
3/28/2016	Hot Topics Course	IUB - Cyberinfrastructure Building	Presented to students various AVL displays.	22
3/28/2016	R: Intermediate Scripts for Humanists	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	10
3/31/2016	Pittsburgh Supercomputing Center: Bridges	XSEDE On-line Workshop	IU hosted one of the virtual sites to introduce this new XSEDE resource.	
4/1/2016	High school outreach	IUB - Cyberinfrastructure Building	High school students from Northern Indiana to see the CIB	6
4/1/2016	Introduction to Scalar	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	1
4/5/2016	Leveraging the Risk Management Framework	IT Training Course IUB / Main Library	A course detailing the NIST based risk management framework UITS has developed for HIPAA/FISMA compliance	50
4/5/2016	Z637 Class presentations and tour	IUB - Cyberinfrastructure Building	Visualization facilities tour and student presentations.	16
4/5/2016	Lustre.Org: A Community Resource	Lustre Users Group 2016 / Online	Presented update on additions and improvements.	180
4/5/2016	S4ES: Introductory Parallel Programming Workshop	IUPUI - Informatics & Communications Technology Complex IT121	Introductory Parallel Programming Workshop	2
4/7/2016	S4ES: Introductory Parallel Programming Workshop	IUB Wells Libray W 144	Introductory Parallel Programming Workshop	6
4/8/2016	RT services, systems, and tools	IUPUI Research Day / IUPUI	Presentations about UITS RT	25
4/8/2016	3D Object Creation with Photogrammetry	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	7
4/11/2016	AVL Tour for IUPUI College Mentors Program	IUPUI - Informatics & Communications Technology Complex AVL	For the College Mentors program and their "Little Buddies" from local elementary schools.	36
4/12/2016	AVL Tour for Computer Science Students from Pike HS	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour of AVL systems and projects	17
4/13/2016	AVL Tour for Washington HS students	IUPUI - Informatics & Communications Technology Complex AVL	Formal tour of AVL systems and projects.	76
4/14/2016	AVL tour for I101 Students	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour about AVL systems and projects	23
4/15/2016	AVL Tour for iDEW	IUPUI - Informatics & Communications Technology Complex AVL	iDEW Program students are guests of the SOIC department and this was a formal AVL facilities tour lead by AVL staff.	26
4/15/2016	AVL Tour for Anna Lynch and family	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour for some colleagues and their family.	7
4/15/2016	360 Degree Panoramas and Open Source 3D Viewer	IUB - Herman B Wells Library	CyberDH Spring Workshop Series	9
4/18/2016	AVL Tours for Professor Zheng Media Class	IUPUI - Informatics & Communications Technology Complex AVL	Tour and informal discussion and demos led by AVL staff	23
4/19/2016	S4ES: Intermediate Parallel Programming Workshop	IUPUI - Informatics & Communications Technology Complex IT121	Intermediate Parallel Programming Workshop	1
4/21/2016	S4ES: Intermediate Parallel Programming Workshop	IUB Wells Library W 144	Intermediate Parallel Programming Workshop	2
4/22/2016	Scripted deployment of Globus Server using Ansible	GlobusWorld 2016 / Chicago IL	Presentation on deploying the Globus file transfer server using Ansible developed to improve the ease of building an XSEDE-like cluster.	80

Date	Education, Outreach, and Training Event Title	Conference Name/Location	Description	Total Attendees
4/28/2016	AVL Tour for Tech 104	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour lead by AVL staff.	18
4/29/2016	Beginning Text Analysis with R	IUB - Herman B Wells Library	Workshop in Methods	25
5/2/2016	AVL Tour for TECH104 Purdue Engineering Freshmen	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour lead by AVL staff.	12
5/11/2016	AVL Tour for FO AP	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour lead by AVL staff.	7
6/1/2016	pti/rt talk			--
6/8/2016	RT services, systems, and tools	IU Libraries Foundation Reception / IUB - Herman B Wells Library	Demonstrated 3D and presented UITS RT services	3
6/9/2016	I-Light Conference 2016	Ivy Tech Conference Center Indianapolis	I-Light Annual Member Conference	45
6/12/2016	Resource Management from HPC to the Cloud	Accelerated Data and Computing (ADAC) Workshop / Lugano Switzerland	Jetstream presentation to an international audience	55
6/13/2016	HILT Open House tours	HILT Open House tours	HILT Open House tours	49
6/14/2016	Cybersecurity Talk	MINI University / IUB - IMU Georgian Room	Introduced Fred Cate and hosted room for Mini University	90
6/15/2016	AVL Tour for SOIC Workshop Students June 15th	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour lead by AVL staff.	38
6/15/2016	PLOW Digital HoloLens discussion and AVL Tour	IUPUI - Informatics & Communications Technology Complex AVL	Informal tour of AVL systems and projects	4
6/16/2016	Educational workshop for jr. high and high school students	Ready, Set, Robots! Advanced Summer Camp / IUB - Cyberinfrastructure Building	Taught campers about computer programming using Lego Mindstorm robots	16
6/16/2016	Robot Grand Challenge	Ready, Set, Robots! Advanced Summer Camp / IUB - Cyberinfrastructure Building	Presentations by campers to family and friends	40
6/16/2016	Tour of Data Center	Ready, Set, Robots! Advanced Summer Camp / IUB - Cyberinfrastructure Building	Took campers, friends, and family on tour after camp.	15
6/22/2016	AVL tour for NEST students	IUPUI - Informatics & Communications Technology Complex AVL	Formal tour of AVL facilities.	35
6/24/2016	Hosted various cybersecurity topics	Cybersecurity Matters Camp / IUB - Cyberinfrastructure Building	Camp for high school students	21
6/24/2016	Educational Activity	Minority Engineering Advancement Program (MEAP) / IUPUI	Tour of AVL facilities for three groups of students in the MEAP camp.	55
6/24/2016	Educational Activity	Minority Engineering Advancement Program (MEAP) / IUPUI	Lasers and fiber optics activities in STEM.	20
6/24/2016	R for Humanists: A Git Repository and Tutorial Approach	Keystone DH / Pittsburgh, PA	Presentation	30
6/25/2016	Hosted various hackathon topics	Galaxy Community Conference Hackathon Day 1 / IUB - Herman B Wells Library	CESG and NCGAS organized the hackathon for GCC16	67
6/26/2016	Hosted various hackathon topics	Galaxy Community Conference Hackathon Day 2 / IUB - Herman B Wells Library	CESG and NCGAS organized the hackathon for GCC16	46
6/26/2016	Hosted various Training topics	Galaxy Community Conference Training Day 1 / IUB - Herman B Wells Library	CESG and NCGAS organized the training for GCC16	137

Date	Education, Outreach, and Training Event Title	Conference Name/Location	Description	Total Attendees
6/27/2016	Hosted various Training topics	Galaxy Community Conference Training Day 2 / IUB - Herman B Wells Library	CESG and NCGAS organized the training for GCC16	144
6/28/2016	Hosted entire conference related to the galaxy project	Galaxy Community Conference / IUB - Indiana Memorial Union	CESG and NCGAS organized this major conference	178
6/28/2016	Resilience a Breach-First Approach to Cybersecurity	12th AMC Security and Privacy Conference / Chapel Hill NC	CACR led a panel that introduced the emerging concept of resilience to a medical audience	30
6/28/2016	StarDock: shipping customized computing environments to the data	SPIE2016 Conference / Edinburgh Scotland	The SciAPT SCA team presented a talk about a prototype research project using Docker and SCA	40
6/28/2016	ImageX: New and Improved Image Explorer for Astronomical Images and Beyond	SPIE2016 Conference / Edinburgh Scotland	The SciAPT SCA team presented a poster	50
6/29/2016	Hosted kick-off for NSF-funded project	Science Gateways Community Institute Kick-off / IUB - Cyberinfrastructure Building	NSF funded project kick-off meeting	10
6/29/2016	SOIC Game Camp AVL Visit	SOIC Game Camp / IUPUI AVL	This was a visit by the SOIC game camp students to try out the POPcon VR system that IU students built in the lab.	35
6/29/2016	IQ-Wall paper presentation at Ed Media 2016 conference	Ed Media 2016 / Vancouver Canada	IQ-Wall paper presentation at Ed Media 2016 conference	6
6/29/2016	Trident: Scalable Compute Archive - Workflows Visualization and Analysis	SPIE2016 Conference / Edinburgh Scotland	The SciAPT SCA team presented a talk about the evolution of the project from ODI to today.	40
6/29/2016	Galaxy security practices in an age of clinical data for point of care services	Galaxy Community Conference / IUB - Alumni Hall IMU	Talk to users and admins to keep security in mind when using and deploying Galaxy	100
6/30/2016	Hosted meeting with various topics	Generic Model Organism Database (GMOD) project meeting / IUB - Indiana Memorial Union	CESG organized this meeting for GMOD.	24
6/30/2016	Fast and flexible big data processing at LSST data rates using existing shared-use hardware	SPIE2016 Conference / Edinburgh Scotland	The SciAPT SCA team's collaborator Ralf Kotulla presented a talk about RabbitQR an astronomical pipeline.	40
			Total Attendees:	**7,705**